Detecting Location Spoofing in Social Media: Initial Investigations of an Emerging Issue in Geospatial Big Data

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Bo Zhao

Graduate Program in Geography

The Ohio State University

2015

Dissertation Committee:

Daniel Z. Sui, Advisor

Harvey Miller

Ningchuan Xiao

Copyrighted by

Bo Zhao

2015

Abstract

Location spoofing refers to a variety of emerging online geographic practices that allow users to hide their true geographic locations. The proliferation of location spoofing in recent years has stirred debate about the reliability and convenience of crowd-sourced geographic information and the use of location spoofing as an effective countermeasure to protect individual geo-privacy and national security. However, these polarized views will not contribute to a solid understanding of the complexities of this trend. Even today, we lack a robust method for detecting location spoofing and a holistic understanding about its multifaceted implications. Framing the issue from a critical realist perspective using a hybrid methodology, this dissertation aims to develop a Bayesian time geographic approach for detecting this trend in social media and to contribute to our understanding of this complex phenomenon from multiple perspectives. The empirical results indicate that the proposed approach can successfully detect certain types of location spoofing from millions of geo-tagged tweets. Drawing from the empirical results, I further qualitatively examine motivation for spoofing as well as other generative mechanisms of location spoofing, and discuss its potential social implications. Rather than conveniently dismissing the phenomenon of location spoofing, this dissertation calls on the GIScience community

to tackle this controversial issue head-on, especially when legal decisions or political policies are reached using data from location-based social media. Only then can we promote more effective and trustworthy geographic practices in the age of big data.

Dedication

Dedicated to the students at The Ohio State University

Acknowledgments

Over the past three years, I have received support and encouragement from a great number of individuals. I wish to offer my most heartfelt thanks to the following people. Dr. Daniel Sui has been a mentor, colleague, and friend. We first met during one of the most difficult times of my life, and he was there to listen and to offer guidance. His patience, flexibility, genuine caring, and faith in me during the dissertation process enabled me to earn my Ph.D. He is motivating, encouraging, and enlightening. He has never judged nor pushed even when he knew I was in the midst of juggling priorities. Thank you for the advice and support to pursue research topics about which I am truly passionate. I see the same drive and passion in your own research efforts, and I thank you for letting me do the same.

I am very grateful to the remaining members of my dissertation committee, Dr. Harvey Miller and Dr. Ningchuan Xiao. Their academic support, input and encouragement are greatly appreciated. Moreover, Dr. Ola Ahlqvist served on my exam committee and provided valuable advice on the literature review. I must thank Dr. Morton O'Kelly who allocated a spacious working environment for me in CURA and other logistics.

To Jake Carr, Yapin Liu, Yuxi Zhao, Xiang Chen, and Shaun Fontanella, you have generously shared your time and ideas. I have learned much through our conversations. To Julia Shaw, Young Rae Choi, Samuel Kay, Lili Wang, I am grateful for the countless hours spent to proof read working drafts. To Xining Yang, I have always been able to turn to you when I have needed a sympathetic ear, an honest opinion, or someone to just grab lunch with. To Ying Song, you have been a critical and insightful reviewer of all my draft. You were always there with green tea latte or cookies whenever I got exhausted. To Ruoran Cheng, your selfless help made me concentrate on completing the dissertation.

Without the love and support of my family, this dissertation would never have seen the light of day. I am particularly grateful to my parents and in-laws, who have always inspired me to follow my heart and honor my deepest passions. Thank you. I extend the greatest gratitude to my wife, Yuanyuan Duan, for encouraging me to follow my dreams, even when that has meant making some unimaginable sacrifices. You have been a constant source of strength, joy, and love.

Vita

1998 ……..……..……	NPU Middle School, Xi'an, China
2002……..……..……	B.S. and M.S. Cartography and Geographic Information System, Nanjing University, China
2008……..……..……	Graduate Assistant, Department of City and Regional Planning, University of Florida
2009……..……..……	Research Assistant, Center for Geographic Analysis, Harvard University
2012 to present……..……..	Graduate Associate, Department of Geography, The Ohio State University

Publications

S.J. Kay, B. Zhao and D.Z. Sui (2014) Can Social Media Clear the Air? A Case Study of the Air Pollution Problem. *Professional Geographer*, 351-363.

D.Z. Sui, Z.W. Xu and B. Zhao (2013) Networks in Geography. In R. Alhajj and J. Rokne, *Encyclopedia of Social Network Analysis and Mining (ESNAM)* Berlin, Springer.

D.Z. Sui and B. Zhao (2013) GIS as Media: Geographies of Media and Mediated Geographies and through the GeoWeb. In S.P. Mains, J. Cupples and C. Lukinbeal, *Mediated Geographies and Geographies of Media*. Netherlands, Springer.

Fields of Study

Major Field: Geography

Table of Contents

Abstract ii

Dedication iv

Acknowledgments v

Vita vii

Table of Contents ix

List of Tables xiii

List of Figures xiv

Chapter 1: Introduction 1

1.1 Motivations and contexts 1

1.2 Objectives 4

1.3 Synopsis of the dissertation 5

Chapter 2: Related Work 8

2.1 Definition and relevant vocabulary 8

2.1.1 Case studies: Spoofing as a geographic phenomenon 9

2.1.2 Connotations and denotations 14

2.1.3 Spoofing process and related elements 24

2.1.4 Locational inconsistency 28

2.1.5 Spoofing motivation 30

2.2 Understanding location spoofing based on geographic knowledge 33

2.2.1 Linguistic fallacy 33

2.2.2 Ontology for understanding location spoofing 37

2.2.3 Necessity of geographic knowledge for detecting location spoofing 41

2.3 Time geography 44

2.3.1 Binary models: Space-time prism 45

2.3.2 Frequentist models: Prison/prism break 47

2.3.3 Bayesian models: A potential direction 49

2.4 McLuhan's laws of media 52

2.5 Summary and research framework 54

Chapter 3: Methodology 58

3.1 Crawling and harvesting social media feeds 59

3.2 Detecting fake locational information 64

3.2.1 Prior beliefs: Individual activity 69

3.2.2 Posteriori probability: Human appearance 71

3.3 Profiling spoofing motivation 74

3.4 Summary 77

Chapter 4: Empirical Results 80

4.1 Settings: Data and variables 81

4.1.1 Description of the data 81

4.1.2 Variables for detection 87

4.2 Results 93

4.2.1 Temporal preferences 94

4.2.2 Spatial distribution 97

4.3 Evaluations 103

4.4 Discussion 108

4.4.1 Motivations for spoofing 109

4.4.2 Controversial aspects of location spoofing 123

4.4.3 Social implications of location spoofing 130

4.5 Summary 133

Chapter 5: Conclusions and Future Research 136

5.1 Summary and concluding remarks 136

5.2 Future research 141

5.2.1 Integrating with other detecting approaches 141

5.2.2 Cluster based spoofing detection 143

5.2.3 Further critical inquiries 144

References 146

List of Tables

Table 1. Top 20 sources measured by the number of tweets 86

Table 2. Detected fake geo-tags 93

Table 3. Nearest neighbor index (NNI) analysis results 99

List of Figures[1]

Figure 1. Fake location at Derby Hall 9

Figure 2. Zilla's fakecation in Southeast Asia [Source: (de Born 2014)] 11

Figure 3. GPS spoofing to mislead a superyacht [Source:modified after Zumalt (2013)] 12

Figure 4. Fake locations in profiles to baffle the Iranian authorities 13

Figure 5 Two fictional towns in Ohio: Goblu and Beatosu,based on Southorn (2015) 15

Figure 6. "Easter egg" Android vs. Apple on Google maps [Source:modified after Woollaston (2015)] 17

Figure 7. Appearance frequencies of the terms "spoof" and "spoofing" in Google n-gram database (1899-2008) 18

Figure 8. Geolocation emulcation on Google Chrome 20

Figure 9. Configuration of a location mock on Android 21

Figure 10. LocationFaker on iPhone [Source: (Bhatnagar 2014)] 21

Figure 11. Mobile GPS spoofer to mislead a car 23

Figure 12. Elements of location spoofing under the 5W model (1948) 26

Figure 13. Schramm's (1954) circular model of communication 27

[1] All the listed figures are made by author of this dissertation unless otherwise notified.

Figure 14. The structure of a sign 29

Figure 15. The structure of locational inconsistency 29

Figure 16. The role of the self in interpersonal communication [Source: Burton and Dimbleby (1988)] 32

Figure 17. The "speaking circuit" between two speakers with different understandings [Source: modified after Saussure (1974, 2013)] 35

Figure 18. Circular model of intrapersonal communication [Source: (Fill 1999)] 36

Figure 19. Stratified ontology of critical realism [Source:modified after Mingers (2004)] 39

Figure 20. Time geographic measuring elements 46

Figure 21. All geo-tagged locations of Flickr photos [Source: Gilmer (2015)] 51

Figure 22. McLuhan's laws of media [Source: Sui and Goodchild (2003)] 53

Figure 23. Research framework [Source: modified after Goodchild et al. (1999)] 56

Figure 24. Open-source data harvesting cluster 63

Figure 25. Flow chart for detecting fake locational inforamtion 65

Figure 26. Detecting a fake path via a space-time cone 67

Figure 27. Measuring errors from a mobile device 73

Figure 28. A real location spoofing case in New York city 78

Figure 29. Global distribution of all harvested geo-tagged tweets 82

Figure 30. Harvested geo-tagged tweets around NYC 82

Figure 31. The temporal dimension of the geo-tagged tweets over the continental United States 84

Figure 32. Tweet reposting via Instagram 85

Figure 33. Geo-tagged tweets in a d-d plot 90

Figure 34. Distribution of tweet density 91

Figure 35. Frequencies of posting geo-tagged tweets 95

Figure 36. Temporal patterns of posting geo-tagged tweets thoughout a day 95

Figure 37. Global distributions of fake geo-tagged tweets 98

Figure 38. Fake path in New York city 98

Figure 39. Cumulative frequencies of nearest neighbor distances 102

Figure 40. Cumulative frequencies of nearest neighbor distances of differnt sources ... 102

Figure 41. Frequencies of the time interval between posting consecutive geo-tagged tweets 105

Figure 42. Fake paths of location spoofing over the continental United States 105

Figure 43. Frequencies of distance between two consecutive geo-tagged tweets 106

Figure 44. Location spoofing apps for Android phones 111

Figure 45. GPS jammers/spoofers in Google Shop 112

Figure 46. Location spoofing for fun 113

Figure 47. Location spoofing during a layover.....115

Figure 48. Location spoofing by a dug dealer 115

Figure 49. Trilateration used on social media.....118

Figure 50. Location shift.....118

Figure 51. Location spoofing as an advertising strategy 120

Figure 52. Controversial geo-tags dent by Taliban dpokesman..... 122

Figure 53. Representation of locational uncertainty 125

Figure 54. Analysis of the posting process of a tweet 126

Figure 55. Interview text with a leftist from New Jersey 128

Chapter 1: Introduction

1.1 Motivations and contexts

Location spoofing, referring to the act of intentionally falsifying one's true location, has been a long-standing but often ignored occurrence. Perhaps, one of the earliest examples can be traced back to the Trojan War in ancient Greece. By camouflaging the location of the battlefield (outside the Trojan city) by presenting a giant wooden horse and numerous pulling-down tents, the Greek army appeared to flee the Trojan city, but in fact, the troops were still outside the city waiting to attack. During the American Revolution War, George Washington also used the similar locational deception tactics in the battle of Trenton – by lighting numerous campfires all throughout the camp to camouflaging it from the British, which allowed Washington's army to slipped away and prepared for a surprise blow at Princeton. Another example of spoofing as a military strategy took place during the Second World War when Operation Fortitude successfully influenced the Nazi army to attack fake field armies at Edinburgh and Pas-de-Calais rather than the real ones in Normandy. On a lighter note, in modern society, the movie industry consistently employs location spoofing

- a technique commonly known as "geopiracy"[2] - to create sets which bring to life another time and place completely different from the present geographical environment. Most movie fans may recall that set designers for the classic film *Gone with the Wind* (1939) were able to transform Hollywood into the Antebellum South (Vogel et al. 2007). More recently, the Chinese Cyber command (PLA Unit 61398) meticulously falsified its location when conducting cyber-espionage (Mandiant 2013). Moreover, we must increase awareness of the public's vulnerability to GPS spoofing/jamming techniques used to disrupt global financial transactions (Economists 2013), commit economic crimes in China (Guo and Jiang 2014), dangerously misdirect a yacht in the Mediterranean Sea (Zaragoza 2013), interfere with the launching and landing processes of aircrafts at the Newark Airport (Gibbons 2013), and even mislead GPS guided missiles or drones (Koerner 2014). Shockingly, it is technically feasible for terrorists to remotely hijack civilian airplanes. We can no longer afford to underestimate or ignore these new capabilities due to their potential devastating consequences if used with malicious intent.

Even though several terms with slightly different connotations frequently appear in the literature, such as location obfuscation, location obscuration, and GPS spoofing/jamming, this critically important issue has not received the attention it deserves from the GIScience community. Especially, as location-based social media have become increasingly

[2] Another major strand of literature (Wainwright, 2012) uses the term "geopiracy" to offer a postcolonial critique of human geography today.

integrated into a variety of our activities during the past five years (Sui and Goodchild 2011, Batty et al. 2012), we can no longer simply dismiss location spoofing as merely a nuisance. With the growing popularity of social media, we have also witnessed a significant increase in location spoofing for a variety of purposes by large corporations, government/military agencies and individuals (CCTV 2013, BBC 2014)[3]. Yet, even though research efforts have been devoted to exploring these phenomena, which include other types of spoofing (e.g., identity, IP address or Email etc.), geographies of social media (Papacharissi 2009, Kitchin et al. 2013), and the poor data quality and uncertainty of geographic information (Goodchild 1993, Zhang and Goodchild 2002), GIS scientists have largely ignored this phenomenon as a critical research topic. This neglect may stem from the fact that society tends to take geo-tags or latitude/longitude coordinates for granted as the actual locations (Java et al. 2009) (Cheng, Caverlee and Lee 2010) (Stefanidis, Crooks and Radzikowski 2013) rather than critically examining the credibility of given location information (Crampton et al. 2013). Our knowledge about location spoofing remains fragmented and very limited. As a result, the GIScience community still has a poor understanding of its broader implications.

[3] CCTV refers to the China Central Television in this dissertation.

1.2 Objectives

In this dissertation, I will take on this challenge by conducting the first study of location spoofing. I hope to achieve the following objectives:

(1) Defining location spoofing and relevant vocabulary. Conceptually, location spoofing is recognized as a typical information transmission in space-time. Its transmission process, constituent elements, and defining components have typically been studied through mainstream communication models. However, I will broaden this research by utilizing an interdisciplinary perspective drawing from geography, linguistics, communication sciences, psychology, and criminology, to investigate the connotations and denotations of location spoofing.

(2) Proposing a quantitative method to detect location spoofing using geographic knowledge and a Bayesian time geographic approach. In light of volunteered geographic information (VGI) and geospatial big data quality assessment, the proposed method fundamentally relies on human constraint in a geographic context to accurately sift out spoofing information. To test the utility and efficiency of this method, I will implement an empirical study to sift location spoofing from millions of geo-tagged tweets.

(3) Interpreting the hidden motivations of location spoofers through online observation and auxiliary interviews. Additionally, I will critically examine some controversial aspects

of this spoofing phenomenon, including the locational uncertainty, software design dilemmas, and computer malfunctions.

(4) Discussing the implications of location spoofing for individuals and society as a whole as they relate to McLuhan's laws of media. I will, thus, encourage a critical and holistic perspective with regard to this spoofing phenomenon. By doing so, I plan to raise awareness rather than underestimating or simply ignoring its existence, which has occurred in the past.

1.3 Synopsis of the dissertation

This dissertation is organized as follows:

Chapter 1 introduces the motivation, general context, challenges and opportunities inherent in the world of geospatial big data and location-based social media. Subsequently, the research objectives are outlined, followed by a synopsis of this dissertation.

Chapter 2 is a review of the recent academic literature related to the topic. Through several typical cases, this chapter defines the term location spoofing and other relevant vocabulary. Resting upon a critical realist perspective of this complicated phenomenon, the critical role of geographic knowledge in understanding location spoofing is emphasized in this chapter. After laying a foundation for the importance of geographic knowledge, this chapter

continues to review the relevant time geography literature, particularly its theoretical background, and binary and frequentist models, and suggests a potential direction for future research of time geography via a Bayesian approach. To better understand the social implications of location spoofing, I relate spoofing phenomenon to McLuhan's laws of media. This chapter concludes by presenting an overarching GIScience research framework for conducting future location spoofing studies.

Chapter 3 proposes a hybrid research method for studying location spoofing, the goal of which is three-fold, including harvesting social media feeds, detecting fake location information, and analyzing the motivations for spoofing. Specifically, this chapter introduces how to assemble a data crawling cluster using open-source hardware. Then, the specific procedure of harvesting large-scale social media feeds using this assembled cluster is proposed. Later this chapter includes a discussion of how to screen fake locational information by utilizing a Bayesian time geographic approach. Having detected the fake locational information, this chapter then proposes a qualitative method for investigating hidden motivations for spoofing through online observation as well as auxiliary interviews. To sum up, this chapter illustrates the detecting process.

Chapter 4 includes tests of the proposed detecting method via empirical study. This chapter begins with a description of the research data (a large-scale data set containing millions of geo-tagged tweets) and variables. After applying the detecting approach for this empirical

study, the rest of this chapter is a detailed discussion of the results and an unveiling of the statistical, spatial-temporal characteristics of this spoofing phenomenon. Also, it includes an evaluation of this approach to show its limitations as well as possible directions for further improvements. Moreover, an attempt is made to examine the multifaceted reasons why spoofers falsify their locational information, and discuss its controversial aspects. Furthermore, this chapter examines the broader implications of location spoofing under the framework of McLuhan's laws of media.

Chapter 5 concludes this dissertation, and proposes three potential directions for future research, including using a cluster to speed up the detecting process, integrating the use of crowd-sourcing or similar approaches with the main proposed geographic approach to discover more location spoofing cases, and employing further critical inquiries regarding the paradigm shift to commutative rationality.

Chapter 2: Related Work

In this chapter, I review recent academic literature related to location spoofing. Taking an interdisciplinary perspective, I define the term "location spoofing" and some other related terms. Moreover, this chapter develops a critical realist argument for the necessity of using geographic data to detect location spoofing. In addition to the geographic approach, I will discuss time geography as a well-developed quantitative geographic framework. Hence, based on a review of the three phases of recent development in time geography, I explore the possibility of employing a Bayesian time geographic approach to detect location spoofing. We must also examine the social implications of this phenomenon according to McLuhan's Laws of Media.

2.1 Definition and relevant vocabulary

Not surprisingly, even if you come across the term "location spoofing" for the first time, you may still be able to name a few apt examples of how people misuse locational information, but upon further reflection, its exact meaning might still baffle you. Therefore, before formally defining this term, I would like to begin by introducing several examples

of how this phenomenon closely affects our daily lives and contemporary society.

2.1.1 Case studies: Spoofing as a geographic phenomenon

- Bob's little white lie

Bob is a Ph.D. candidate, studying at the Ohio State University. He and his fiancée planed to marry in the next couple of weeks. Because of his job as a teaching assistant and the writing of his dissertation, he has not been able to see his fiancée as much as he would like. In order to alleviate her anxiety, Bob frequently sent geo-tagged tweets to keep in touch when he was working very late at his office. These tweets proved to her that Bob was working at his office rather than hanging out all night in a bar.

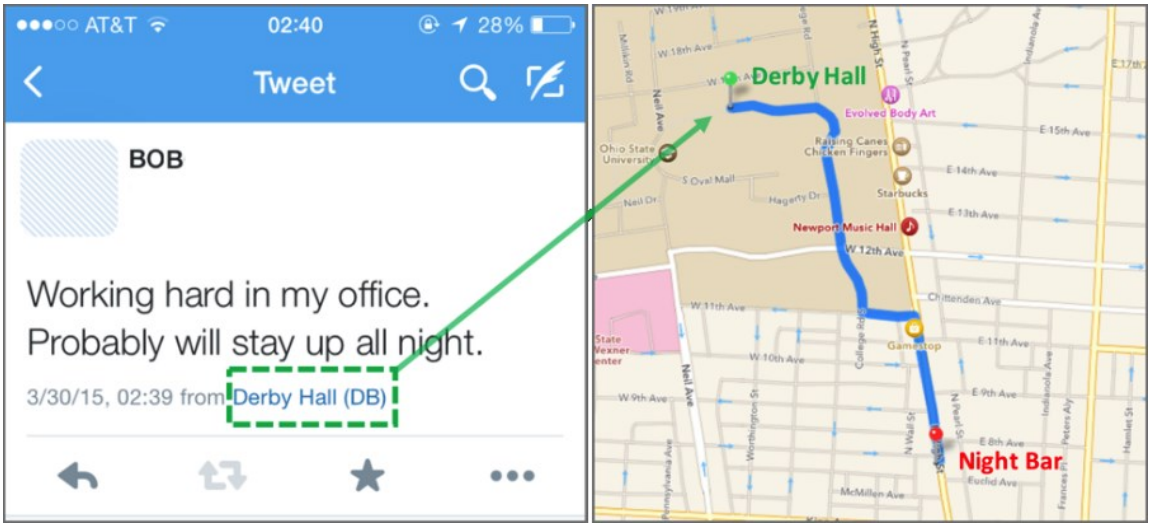

Figure 1. Fake location at Derby Hall

Bob's lab partners were well aware of his coming wedding, so they all insisted on

organizing a bachelor party at one of the bars on High Street. This request embarrassed Bob. He really wanted to hang out with his friends, but he knew his fiancée would not understand and worry about him.

To solve this problem, Bob decided to tell a white lie about where he would be that night. Using a mobile app called "Location Faker", Bob made it seem that he would be working all night at his office by sending his fiancée a tweet that came with a geo-tag at Derby Hall where Bob's office was located. In fact, he was already hanging out with his friends at a bar on High Street. After his fiancée viewed his geo-tagged tweet, she believed that Bob was working hard at the office. She sighed and wondered if he was working too hard as she fell into a peaceful sleep.

- Fakecation

Zilla van den Born, a 25-year-old Dutch girl, went on a five-week backpacking "Fakecation" to several famous tourist destinations in Southeast Asia (see Figure 2[c]). During a fakecation, people do not actually visit a place, but can provide their audience with the illusion that they have been there.

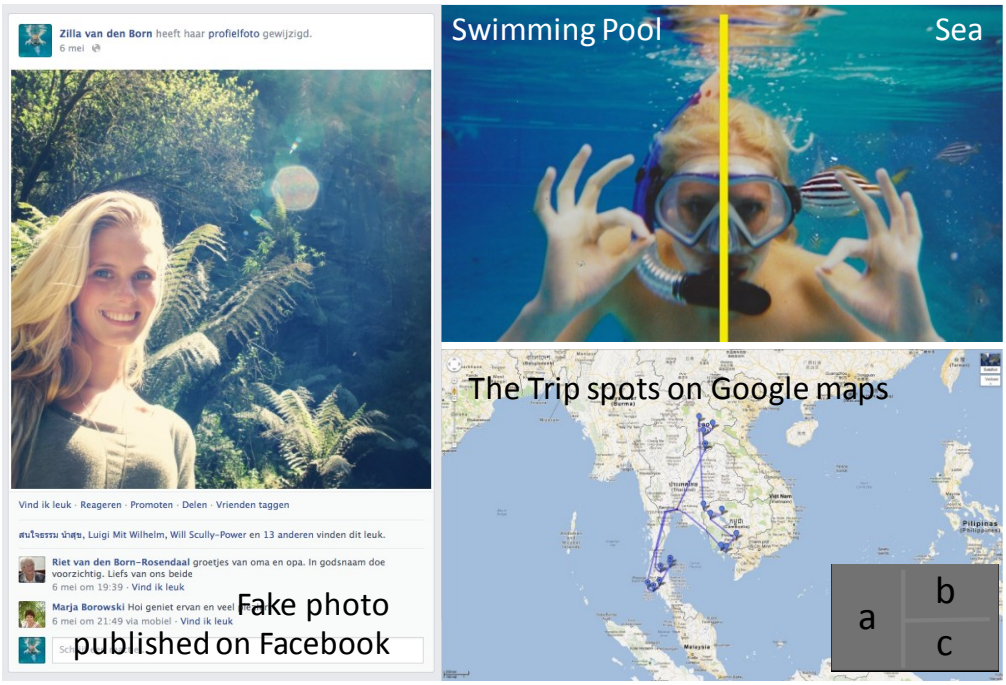
Zilla van den Born heeft haar profielfoto gewijzigd.
6 mei
Vind ik leuk · Reageren · Promoten · Delen · Vrienden taggen
สนใจธรรม ท่าสุข, Luigi Mit Wilhelm, Will Scully-Power en 13 anderen vinden dit leuk.
Riet van den Born-Rosendaal groetjes van oma en opa. In godsnaam doe voorzichtig. Liefs van ons beide
6 mei om 19:39 · Vind ik leuk
Marja Borowski Hoi geniet ervan en veel
6 mei om 21:49 via mobiel · Vind ik leuk
Fake photo published on Facebook
Swimming Pool
Sea
The Trip spots on Google maps
a
b
c

Figure 2. Zilla's fakecation in Southeast Asia [Source: (de Born 2014)]

In reality, Zilla never left Amsterdam but stayed in her apartment to flood her Facebook page (Figure 2[a]) with 42 days of posts with fake pictures of her snorkeling (Figure 2[b]), traveling in traditional Thai tuk-tuks, visiting Buddhist temples, and trying out authentic Southeast Asian cuisine. All of the photos were taken in her room, a local swimming pool or some locations around Amsterdam. Of course, these photos were modified with Photoshop. For example, the left half of Figure 2(b) shows the original photo shot in the swimming pool, whereas the right part shows the modified photo, giving the illusion that Zilla was snorkeling with tropical fishes in some vast, exotic sea. Creating this fakecation helped Zilla to prove how common and easy it is to distort reality. She said, "Everybody knows that pictures of models are manipulated. But we often overlook the fact that we

manipulate reality also in our own lives."

- Misleading a yacht

A radio navigation research team from The University of Texas, Austin successfully spoofed a 213-foot super yacht to alter its planned course when traveling from Monaco to Rhodes on the Mediterranean Sea (Zumalt 2013).

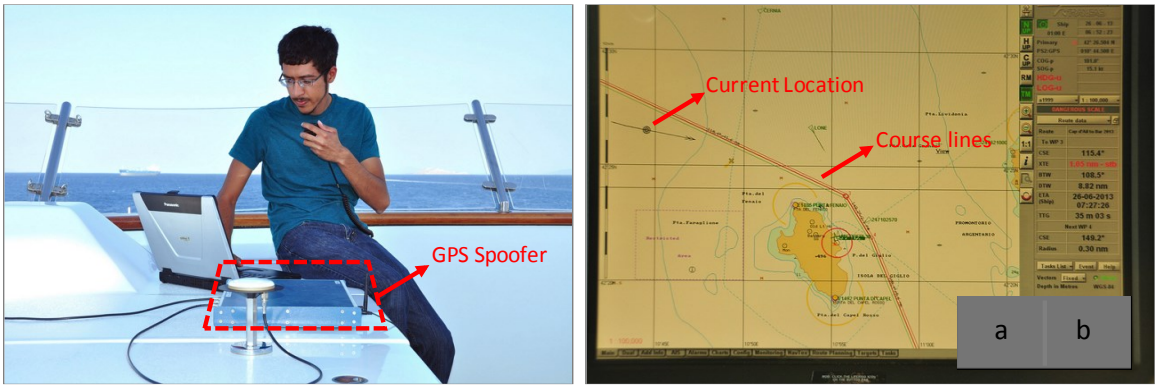

Figure 3. GPS spoofing to mislead a superyacht [Source:modified after Zumalt (2013)]

As shown in Figure 3(a), an operator, on the yacht's upper deck, broadcasted fake GPS signals from his spoofing device - a blue box about the size of a briefcase - toward the GPS antennas of the ship. Consequently, the yacht's GPS readings on the ECDIS (Electronic Chart Display and Information System, see Figure 3[b]) were changed – displaying that the yacht was drifting off the course. This spoofing device caused the yacht's navigation system to spontaneously modify its course. The planned course was determined by a set of waypoints connected by straight red lines, whereas the yacht's current location and direction of travel, indicated by the black target and arrow respectively, was far away from

the original course.

- Occupying Tehran

In reaction to the 2009 presidential election in Iran, dissatisfied voters who believed the election was hijacked by the Green Movement[4], organized several protests to take place around major squares in Tehran as well as online. Some of these online demonstrations were supported by Twitter users who established fake locations in Tehran through spoofing, rather than actually traveling there.

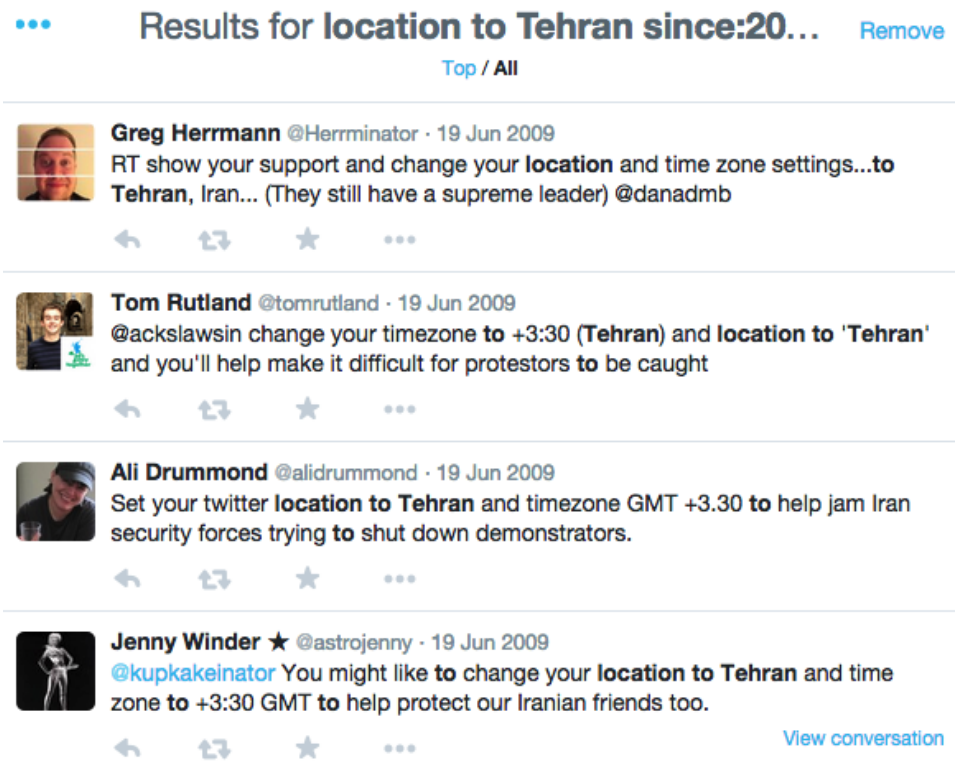
Results for location to Tehran since:20...
Remove
Top / All
Greg Herrmann @Herrminator · 19 Jun 2009
RT show your support and change your location and time zone settings...to Tehran, Iran... (They still have a supreme leader) @danadmb
Tom Rutland @tomrutland · 19 Jun 2009
@ackslawsin change your timezone to +3:30 (Tehran) and location to 'Tehran' and you'll help make it difficult for protestors to be caught
Ali Drummond @alidrummond · 19 Jun 2009
Set your twitter location to Tehran and timezone GMT +3.30 to help jam Iran security forces trying to shut down demonstrators.
Jenny Winder ★ @astrojenny · 19 Jun 2009
@kupkakeinator You might like to change your location to Tehran and time zone to +3:30 GMT to help protect our Iranian friends too.
View conversation

Figure 4. Fake locations in profiles to baffle the Iranian authorities

[4] Green Movement is a political force in Iran calling for fundamental transform in the authoritarian government's structure.

The government of Iran regularly monitors all activities on social media (Ansari 2012). During the campaign, social network sites were suddenly blocked and online political activity became the target for harsh criticism and reprisals from the government. In order to prevent this surveillance as well as protect the online protestors, many internationally based Green Movement supporters spread disinformation over Twitter to mislead foreign observers. Foreign supporters who were not in Iran decided to set their online locations to Tehran[5] in order to protect those who were tweeting from Tehran (see Figure 4). This strategy may have helped some Iranian opposition leaders avoid persecution, but it also made it impossible to understand the real impacts of Twitter on the situation of the demonstrations on the ground.

2.1.2 Connotations and denotations

The term "spoof" first appeared in 1889 when Arthur Roberts (1852-1933), a British comedian, invented the word to describe a hoaxing and nonsensical trait[6] of a poker game. Similarly, Merriam-Webster Dictionary indicates that "spoof", as a noun, means "hoax, deception, or a light humorous parody". When used as a verb, it means, "to copy or

[5] Refer to https://twitter.com/search?f=realtime&q=location%20to%20Tehran%20since%3A2009-06-16%20until%3A2009-06-20

[6] The rule of the spoofing game can be found at https://homepages.warwick.ac.uk/~phscz/misc/spoof.html. Also, another newer version can be found at http://allaboutfunandgames.com/spoof-old-card-game-by-milton-bradley

exaggerate (someone or something) in order to make fun of" or "to cause to believe what is untrue". In the Collins dictionary, "spoof" as a noun, indicates "a joke, a deception, a light parody or satire"; or used as a verb, means, "To fool, deceive" or "satirize in a playful, amiable manner". Notably, in the above-mentioned definitions, "spoofing" refers to the gerund or present participle of "spoof" without any specific connotation. However, in fact, the word "spoofing" is often used as jargon in various professions. Finance, for example, uses it to describe a form of market manipulation, such as pump and dump (Henning, 2014), and in biotechnology, spoofing denotes a set of techniques for counteracting genetic surveillance (Harvey et al., 2013).

Figure 5 Two fictional towns in Ohio: Goblu and Beatosu,based on Southorn (2015)

In addition, spoofing, as a geographic phenomenon with deceptive traits, often appears in

classic cartography and modern geo-visualization. Monmonier (1996, 1) listed various methods for lying/spoofing with maps, and argued that the lie in maps is an inescapable cartographic paradox - "to present a useful and truthful picture, an accurate map must tell white lies". Interestingly, Monmonier, even consider map generalization to be a type of spoofing, mainly because mapmaker musts adjust (generalize or simplify) the representation of geographic phenomena in order to create a suitable and useful maps of different scales and for various purposes. The process of map generalization requires weighing the geographic ground truth and the ultimate utility of a map. For the purpose of protecting geo-privacy, modern geospatial techniques, such as location obfuscation, necessarily "spoofs" one's location slightly to avoid exposing the true location. Moreover, spoofing is used to prevent copyright infringements (e.g., editing the competition, including trap streets) or making "Easter eggs" simply for fun. Figure 5 shows a part of the 1978 - 1979 official state of Michigan map, drawn by a Michigan alumnus, displaying two non-existent Ohio towns "Beatosu" and "Goblu" in Fulton and Lucas Counties, respectively. These two fictional names refer to the slogan of University of Michigan fans "Go Blue!" and a reference to their archrival from the Ohio State University (OSU). Another example of an extra is the "Easter egg" that Google placed in Pakistan (33°30'52.5"N 73°03'33.2"E) on April 24th, 2015 (Willis 2014). This egg, which looks unmistakably like the Android logo, is shown urinating on the Apple mascot (see Figure 6).

Putting aside the debate on who placed this landmark - either Google itself or other individuals, its intent was not very kind. So this improper landmark was promptly deleted.

Figure 6. “Easter egg” Android vs. Apple on Google maps [Source:modified after Woollaston (2015)]

Spoofing, generally referring to a means of tricking computers or their users by hiding one’s identity or faking the identity of another user on the Internet, has been recognized as an effective anti-piracy measure (Chartrand 2004, Murray 2003). According to the frequency of appearance of the words “spoof” and “spoofing” from 1899 to 2008 on Google n-grams (Figure 7), while the use of “spoof” steadily increased, the use of the term “spoofing” sharply increased from 1990 to 2008, which was actually the period when the Internet became popular. The obvious relationship between the popularity of spoofing and

the development of the Internet may suggest that the spoofing phenomenon is a characteristic feature of the Internet per se.

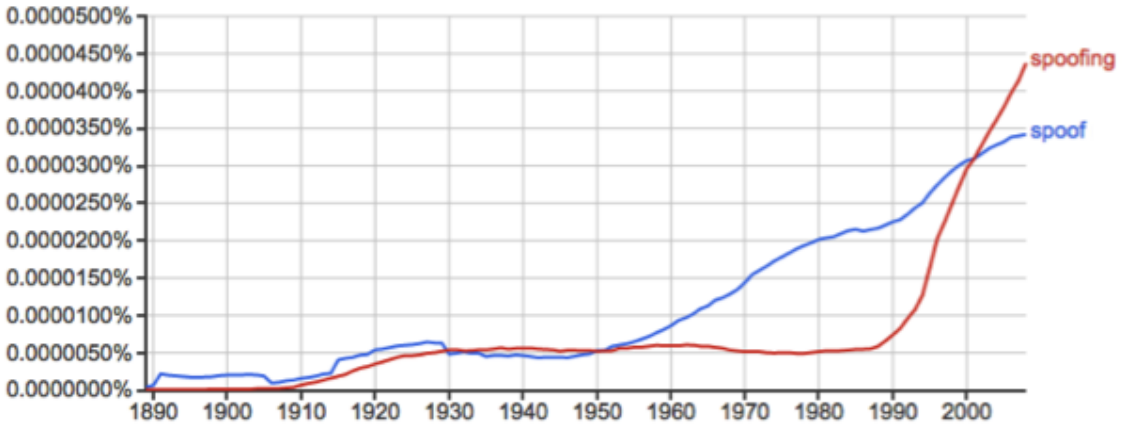

Figure 7. Appearance frequencies of the terms "spoof" and "spoofing" in Google n-gram database (1899-2008)

There are many different types of spoofing on the Internet, including E-mail spoofing (Pandove, Jindal and Kumar 2010, Dodge Jr, Carver and Ferguson 2007), IP (Tanase 2003), Caller ID (Castiglione, De Prisco and De Santis 2009), content/Web (Felten et al. 1997), and GPS spoofing (Humphreys et al. 2008). For example, E-mail spoofing involves sending messages from a bogus E-mail address or faking the E-mail address of another user. Most E-mail servers have security features that prevent unauthorized users from sending messages. However, spammers often send spam messages from their own SMTP, which allows them to use fake E-mail addresses. Therefore, it is possible to receive E-mail from an address that is not the actual address of the person sending the message. Moreover,

IP spoofing can hide or fake the IP address of a computer, and thereby make it difficult to determine the computer's location. This may cause the server to either crash or become unresponsive to legitimate requests. Lastly, spoofing can be as simple as faking one's online identity. For example, in the online chat room, a participant can pretend to be anyone. On most social media, such as Twitter, Instagram, users can fake their age, gender or location.

Of the lexical morphological similarities, location spoofing can be regarded as a specific type that manipulates "locational information". Nevertheless, unlike E-mail, IP or personal identity, locational information is not one specific entity or even a combination of services, protocols, or contents, but exists in multiple forms. For example, locational information can be the raw coordinates acquired by GPS receivers, the IP values when geo-referencing corresponding countries, cities, or even organizations, the top-level domain (e.g., us, uk, cn, jp, kr, etc.) of an E-mail address, the regional or city code of a telephone number, or the textual place name in a user's profile or other web content. More importantly, according to the Open Systems Interconnection model (OSI)[7], the Internet can be perceived as being made up of seven distinct layers and functions, from the very core to the periphery, including application, presentation, session, transport, network, data-link and physical

[7] The Open Systems Interconnection model (OSI) is a conceptual model that characterizes and standardizes the internal functions of a communication system by partitioning it into abstraction layers.

layers. Hence, location spoofing might include one or a combination of various kinds of spoofing that result in fake locational information.

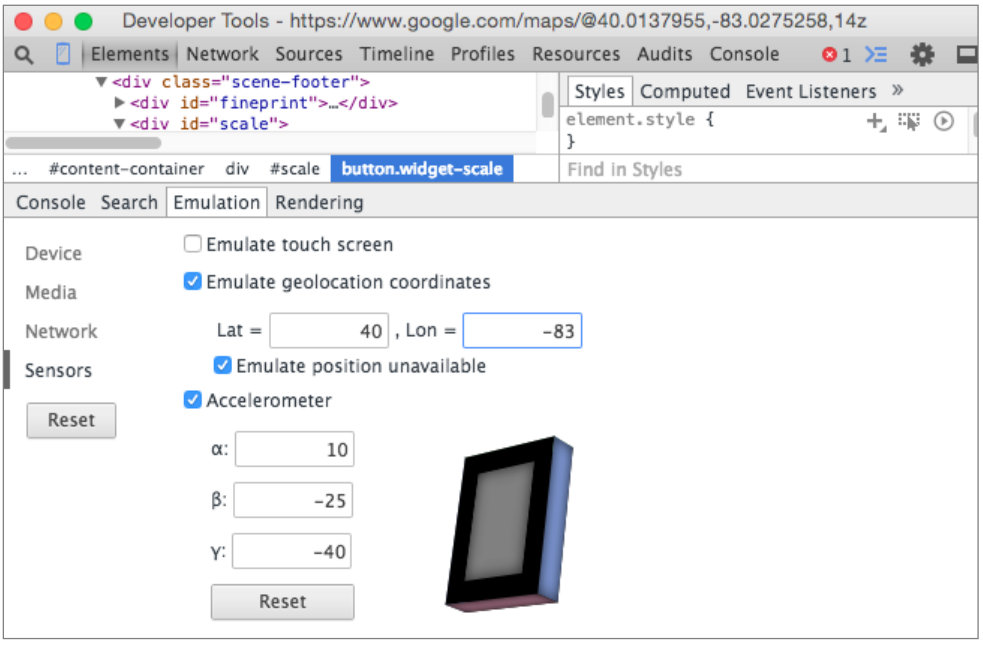

Figure 8. Geolocation emulcation on Google Chrome

To further investigate the spoofing phenomenon, three popular location spoofing techniques are introduced as follows. First, personal computers can more accurately detect one's location via built-in GPS, Wi-Fi networks, and Bluetooth rather than an IP address alone. Most current web browsers, such as Chrome, Firefox, Safari or Internet Explorer, support these built-in positioning plugins, and provide an "HTM5 Geolocation" API to manipulate, debug and even fake locational information. Take Chrome, for example (see Figure 8), users can customize fake coordinates under the "Emulation" tab and open

deceiving social media web sites in Chrome. By employing the method mentioned above, users can easily fake their locational information while posting tweets. After submitting a fake geo-tag to a browser like Chrome, the Twitter server would make it seem that the user is in the fake location. Moreover, most third-party websites do not allow users to attach fake locations to their tweets, but some websites, such as PleaseDontStalkMe.com, give a great deal of latitude to manipulating locational information. On this web site, one can pick any location on a Google Map and then tweet while using this fake geo-tag. Since the Twitter website no longer displays the app name that was used to send that tweet, one's followers on Twitter are less likely to know that one's locational information is spoofed.

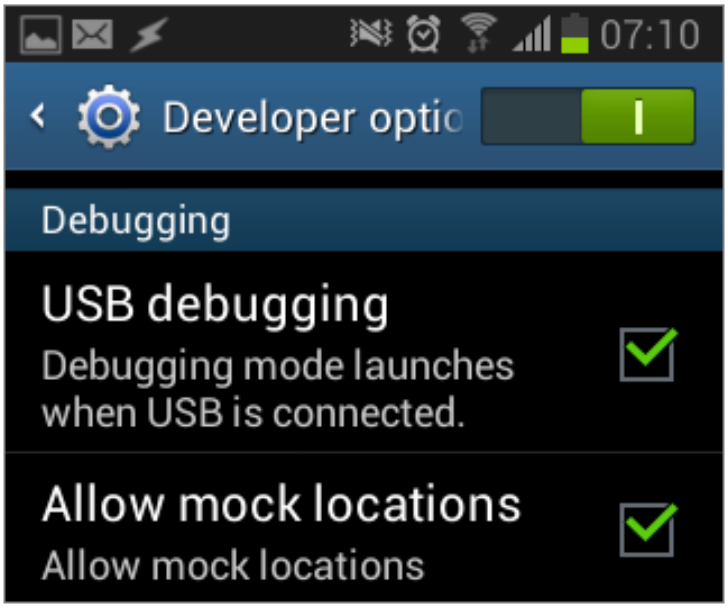

Figure 9. Configuration of a location mock on Android

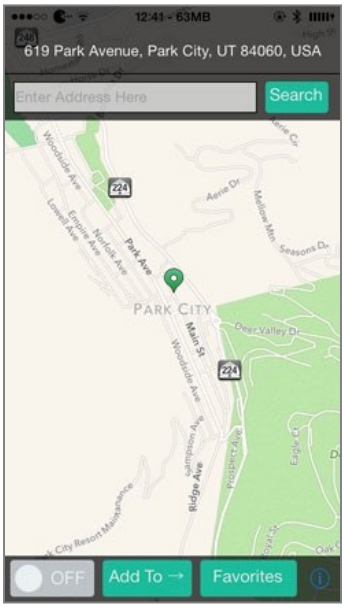

Figure 10. LocationFaker on iPhone [Source: (Bhatnagar 2014)]

Secondly, most mobile devices allow users to modify their locational information to some extent. The function for faking locational information is inherently supported on both Android and iPhone - the two best selling mobile devices. On the Android phone, this function was designed to help the developers debug location-based apps, such as those that determine the geographic outliers, assess the stability in different geographic environments and so on so forth. It is necessary to simulate check-ins or possible routes to avoid physically visiting various places, which saves a lot of time and energy. At the same time, this function can also be utilized for location spoofing. Figure 9 shows the "Allow Mock Locations" option in an Android device. By simply selecting this option, an Android user can customize locational information and cover up the true locational information determined by the built-in GPS sensor. Unlike the Android phone that provides the mock location function, iPhone follows rather strict rules to ensure the accuracy of locational information. As a result, only "jailbroken" iPhones can conceal the built-in GPS sensor. The "Location Faker" app, available in the iPhone Cydia store, allows users to create fake locations (see Figure 10). Thus, iPhone users can falsify the geo-tag to any location by simply choosing a different place on the map screen or inputting the address.

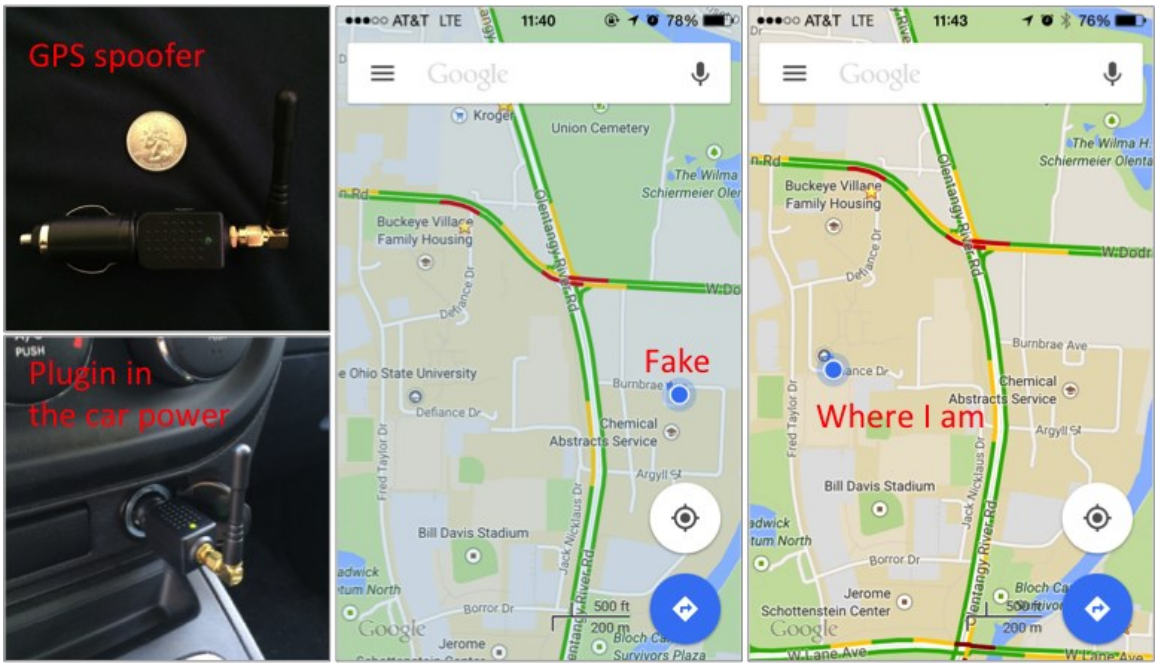

Figure 11. Mobile GPS spoofer to mislead a car

Lastly, a radio frequency (RF) spoofer is often used to deliberately block, spoof, or jam wireless signals, such as GPS and 3G/4G signals. Intentional communication spoofing is aimed at radio signals in an attempt to disrupt them. Specifically, a transmitter, when tuned to the same frequency as the spoofed audience's mobile devices, can override any signal at the receiver. A GPS signal, transmitting on a frequency between 1.5 and 1.6 MHz, is a typical radio signal band that can be misdirected by very simple spoofing devices. As a result, within range, the fake GPS signals can distort the signals of mobile devices, cell towers, civil aviation systems, and other wireless signal resources. Truck drivers often use this GPS spoofer to escape the boss's surveillance. If a GPS spoofer is running near a social media app on a mobile device, it is difficult to accurately geo-tag a location. Figure 11

shows a typical GPS spoofer specifically designed for cars. Based on my tests, when this device was plugged into the car outlet, I found that the locational information acquired by the built-in GPS receiver of my mobile device could transmit as far as three kilometers. One test, for example, showed that when the spoofer was turned on, it displayed the car's location on Google maps (on Burnbrae Avenue) as approximately 2.5 kilometers away from its actual location (on Defiance Drive).

Considering the above-mentioned issues, this dissertation defines location spoofing as an act of intentionally falsify locational information to a place other than one's true geographic location. It most commonly is used in a much narrower sense to indicate the spoofing phenomena on the Internet. Essentially, location spoofing, involves a locational inconsistency between the geographic ground truth and the locational information. Also as it is an intentional act, one must be motivated to consciously choose to create a fake location for spoofing. Overall, location spoofing involves two essential components: the locational inconsistency and the spoofing motivation. Bearing this definition in mind, the next section examines it from the perspective of information transmission, which will explain the spoofing process in greater detail.

2.1.3 Spoofing process and related elements

During the past sixty years, particularly after the Second World War, there were impressive

developments in system science and various communication models were invented. In this section, location spoofing is examined as an information transmission through two influential communication models, Lasswell's 5W model (1948) and Schramm's model (1954).

Lasswell's 5W model (see Figure 12) is named after the five elements of the mode, which are of the same letter "W", that is, Who, (Say) What, (In) Which (Channel), (To) Whom and (With) What (Effects). To interpret location spoofing according to Lasswell's 5W model, we revisit Bob's white lie (see Section 2.1.1) as an example. At the first glance, I notice at least three elements: Bob (the location spoofer), Bob's fiancée (the spoofed audience) and the geo-tag, which states "Derby Hall (DH)" (the fake locational information). These three elements must be linked by some form of channel, or namely, the medium. This would be the cell phone on which Bob composed the geo-tagged tweet, the Internet transmitting the tweet, and the fiancée's cell phone where the tweet was ultimately transmitted. Moreover, we must also consider the effects of this type of process. That is, this fake geo-tag was created to trick Bob's fiancé. Of course, it is critical that she believed it. Overall, a complete information transmission process of location spoofing should include these five elements (see Figure 12).

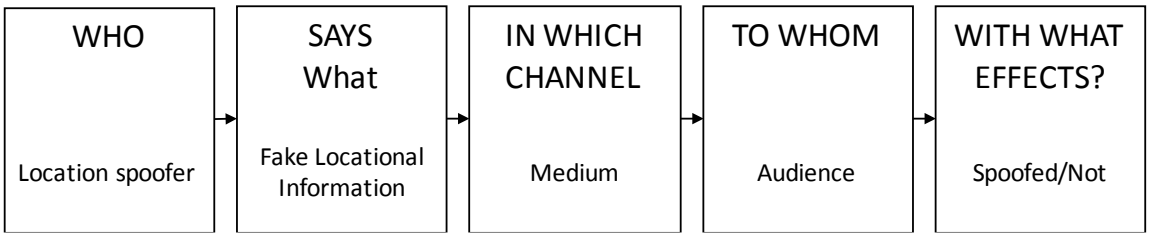

Figure 12. Elements of location spoofing under the 5W model (1948)

(1) Location spoofer (Who): The source or disseminator of the fake locational information, who is responsible for disseminating and processing the information, is an agent who triggers the communicative behavior. This can be either an individual or an organization.

(2) Fake locational information (Message): A set of meaningful signs delivering fake information about the determined location. Here, a sign is defined as the fundamental element of the message. A combination of signs delivers complete meaning.

(3) Medium (Channel): The carrier transfers the message and can be viewed as the linkage of multiple communicative elements. In our lives, a variety of media (e.g., posts, telecom, telephone, TV, Internet, etc.) are used to transmit the fake location information.

(4) Audience (Whom): The information destination, or receiver which the subject/audience (disseminator) wishes to spoof. The disseminator can be affected by reactions or feedback. Likewise, the audience can be either an individual or an organization.

(5) Spoofed/Not (Effects): The reaction from the audience once the fake locational

information is received.

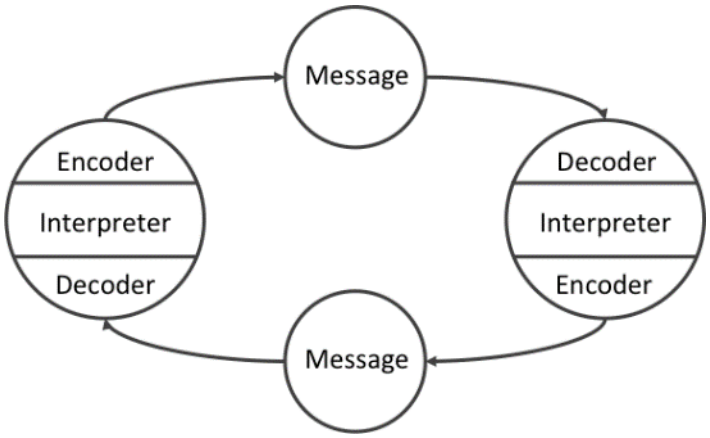

Figure 13. Schramm's (1954) circular model of communication

Of course, due to the complexity and diversity of the information transmission process, the elements are not limited to the five identified above. For example, the message can be divided into "sign" and "meaning", while the audience can be separated into "sender" and "encoder". In addition to analyzing the elements, McLuhan (1964) deemed that the message and the medium (channel) should be combined as an organic whole. It must be noted that, communication is not a one-way street. In order to overcome the limits of a linear model, Schramm (1954) proposed a circular communication model, shown in Figure 13, in which the source is the one sending the message. Encoding denotes the process of transferring a message into a symbolic medium that can be transmitted. Decoding denotes the process of restoring the intended message so that it can be readily understood. Obviously, the receiver is the recipient of the intended message. Feedback is what the

receiver gives back to the source, which enables the source to adjust his or her message to the needs of the audience. Without feedback, it becomes relatively difficult to determine whether the communication has been successful, or to evaluate the impact of the information transmission per se.

2.1.4 Locational inconsistency

Essentially, location spoofing represents the locational inconsistency between the locational information and the geographic ground truth. In reality, one can observe this inconsistency from a piece of "fake" locational information. The major reason it is considered "fake" is due to its inability to accurately represent the geographic ground truth. As far as location-based social media feeds are concerned, the locational information can be in four basic forms: (1) the embedded video, image or audio, which delivers the locational information of the specific information entity. For instance, the photo gives a fake locational impression that the Dutch girl was in Thailand; (2) the textual and semantic geographic information in the content itself. In the sentence "The building is near the campus", the term "campus" functions as a piece of textual geographic information to describe a location, whereas "near" represents the spatial relationship with the given location "campus"; (3) the structured place name listed in a user's profile or other web forms; and (4) the geographic information in the form of longitude/latitude coordinates,

such as geo-tags, check-ins, tracking routings, geo-fencing, etc.

From a linguistic perspective, the observed fake locational information can be referred to as a specific sign that contains meaning. Here, meaning is something other than the sign per se, and it transfers mutually between the sign generator and others who interpret it. Moreover, this sign can be any object, event or entity, whose presence or occurrence indicates the occurrence of something else. For example, the geo-tag "Derby Hall (DB)" was a sign transmitted by Bob to his fiancée, or the black arrow on the monitor was a sign showing the location of the yacht. In the audience's minds, these signs are linked to relevant events. For example, when she read the geo-tag, Bob's fiancée thought he was working late at the office; or by interpreting the movements of the black arrow, the captain thought the yacht had gone off course.

Figure 14. The structure of a sign

Figure 15. The structure of locational inconsistency

According to Saussure's (1974, 2013) linguistic theory, a sign consists of a signifier and the signified. Whilst the signifier acts as a pointing figure, a word, or a sound-image, signified is the meaning or the aspect indicated by the signifier. As shown in Figure 14,

Saussure used a dyadic model of the two objects, a signifier and the signified, to illustrate the internal structure of a sign and deemed that these two objects were inseparable. Depending on the internal structure, we can further understand the concepts of locational (in)consistency, especially the relationship between ground truth as the signified and the locational information as the signifier. As shown in Figure 15, the locational inconsistency essentially indicates a mismatch between these two objects. Rather than directly perceiving the locational inconsistency, the audience can only observe the locational information that encompasses an offset to one's true location. In other words, the geographic ground truth is concealed, which makes the locational inconsistency rather difficult to determine only based on the locational information alone.

2.1.5 Spoofing motivation

In criminology, a crime typically requires the defendant has certain level of motivations for his or her crime. This motivation is usually referred to as "*mens rea* (Latin for "guilty mind"). The standard test for criminal liability states in Latin: "*Actus reus non facit reum nisi mens sit rea*", meaning, "The act is not culpable unless the mind is guilty". In this regard, a motivation is deemed as a precondition for a crime. By the same token, location spoofers must be driven by certain level of motivation as well. Moreover, there are different levels of *mens rea*, and its definition varies between jurisdictions. Particularly the

Model Penal Code (Wechsler et al. 1962) has been highly influential across the United States for clarifying these various levels of culpability. Specifically, there are four levels of *mens rea*, including committing a crime (in descending order) purposely, knowingly, recklessly, and negligently. In order for the defendant to be indicted, he or she must have been found to be in a mental state of recklessness or greater when the crime was committed. Notably, all four levels assume the defendant as a reasonable person[8], who is at least "aware of a substantial and unjustifiable risk that the material element poses or will result from his/her conduct" (Wechsler et al. 1962, 21). Motivations for location spoofing can be also classified into four categories. Therefore, the location spoofer should be of sound mind, and commit the act with negligent or reckless intent. Otherwise, mere locational inconsistency without any intention cannot be considered as location spoofing.

In addition to a legal interpretation, spoofing can be also understood as an intention to share a specific type of geographic information. Coleman, Georgiadou and Labonte (2009) have compiled eleven fundamental motivations for virtually sharing geographic data, including eight positive reasons, including, altruism, professional or personal interest, intellectual stimulation, protection or enhancement of a personal investment, social reward, enhanced personal reputation, creative or independent self-expression, and pride of place; as well as three "negative" ones: mischief, agenda setting and malice and/or criminal intent.

[8] Historically known as reasonable man, this term describes a person who acts with common sense and a good mental state.

In addition, several studies have attempted to understand these motivations though theoretical frameworks, such as the social exchange theory (Mayer and Gavin 2005), social capital theory (Smith et al. 2006), and the power and superiority (Gupta and Govindarajan 2000), and use and gratification theories (Adams 2009). Also, empirical studies have been conducted to interpret knowledge sharing through one or two particular motivations/reasons, such as altruism or reciprocation (Kankanhalli, Tan and Wei 2005), impression management and attribution (Bolino 1999), evaluation apprehension (Bordia, Irmer and Abusah 2006), and social costs (Stasser and Titus 2003).

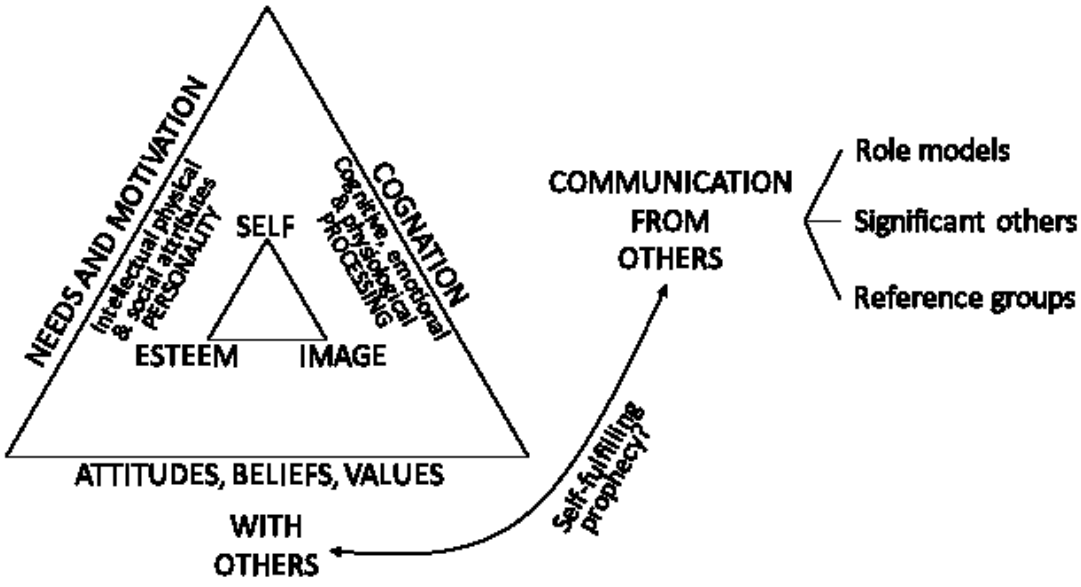

Figure 16. The role of the self in interpersonal communication [Source: Burton and Dimbleby (1988)]

Specifically, Burton and Dimbleby (1988) proposed an interpersonal model that

emphasizes the role of the self in the communication process. As shown in Figure 16, the message flows back and forth between the self (source, the triangle) and others (receiver). This model integrates a triad of self-concept through which the motivation for spoofing can be discovered. In greater detail, one side of the triad represents self-image, which is the way we perceive ourselves. This includes our appearance, personality, background, characteristics, interests and abilities, everything that we portray to others. The other side is our ideal self, indicating how we wish to be in all areas of life, which can be influenced by celebrities, parents, role models, or those with a higher social class. The last category is self-esteem, which shows how we value ourselves. It indicates the appearance and behaviors that are restricted by individual morality and social mores. Overall, these three aspects make up the notion of self-concept. Thus, by linking this model with location spoofing, the self-concept provides a more holistic model for interpreting reasons for transmitting spoofed locational information.

2.2 Understanding location spoofing based on geographic knowledge

2.2.1 Linguistic fallacy

As previously stated, locational inconsistency typically characterizes the mismatch between geographic ground truth (signified) and the locational information (signifier). From a conceptual perspective, locational inconsistency is dependent on the existence of

locational consistency. Because they have a symbiotic relationship, each collapses without the other. Surprisingly, Saussure firmly denied the existence of any consistency between signified and signifier, as he argued: “There is not a natural relation between a word and the object it refers to, nor is there a causal relationship between the inherent properties of the object and the nature of the sign used to denote it” (Saussure 1974, 2013, 114). In other words, the signifier is perpetually floating and unfixed, and the bond between signifier and signified is “essentially arbitrary”. Lacan (1960) extended this theory when he defined this unstable relationship as a “chain of signifiers”. Moreover, most poststructuralist theorists postulate a complete disconnection of the signifier and the signified. By the same token, there is an inevitable disconnect between locational information and geographic ground truth, and any locational information could by no means describe a specific geographic ground truth. Obviously, this idea goes against common sense, and is therefore, not helpful for understanding location spoofing or discovering other scientific knowledge that highly values experiences, evidences and practices.

In fact, this disconnection has become a longstanding issue in linguistics, often referred to as “linguistic fallacy” (Hahn 1957), which is caused by ambiguous terms that eventually results in misunderstandings. Take Bob’s white lie (see Section 2.1.1) for example. Because the geo-tagged “Derby Hall (DH)” indicates a specific place at The Ohio State University, Bob’s fiancée and his colleagues are aware of the link between the displayed

geo-tag and the specific place. However, for those from the University of Nottingham, Derby Hall refers to a building for student housing. Apparently, a term may have different meanings for people in different places and/or in different geographic contexts. In other words, Derby Hall is known by different names, such as, “the geography department”, “154 north oval mall”, etc. Therefore, a specific piece of locational information may signify multiple places depending on context, and a specific place may be signified by multiple pieces of locational information. In this sense, this linguistic fallacy essentially dismantles the binary structure between the locational information and the ground truth, and calls into question any form of geographic representation, including this spoofing phenomenon.

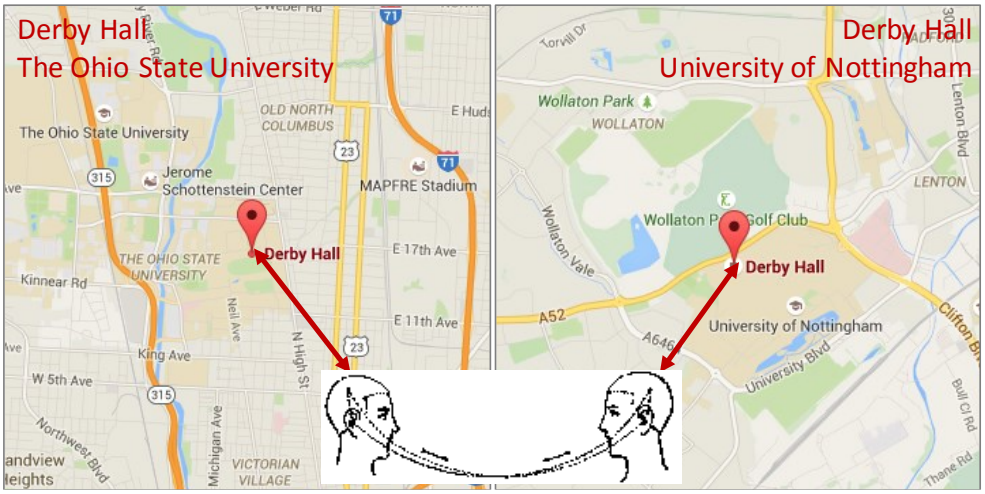

Figure 17. The “speaking circuit” between two speakers with different understandings [Source: modified after Saussure (1974, 2013)]

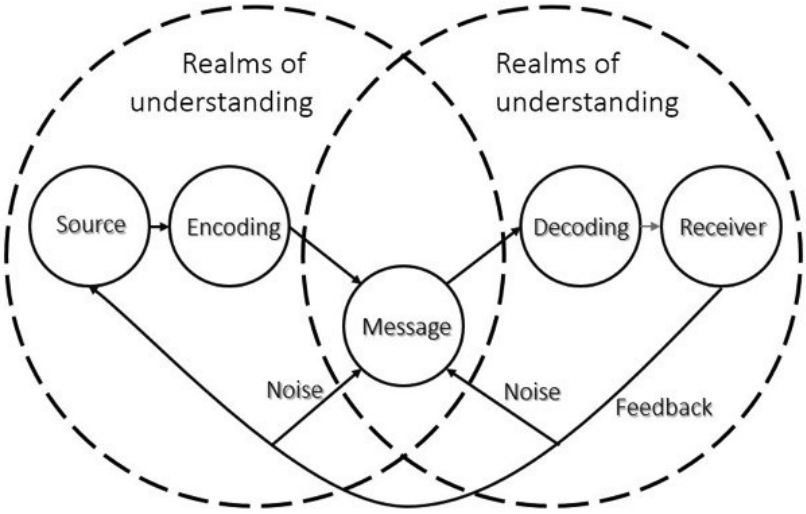

Figure 18. Circular model of intrapersonal communication [Source: (Fill 1999)]

Many efforts have been devoted to dealing with this thorny issue. Saussure argued that a shared system of language could ensure the transmission of the message unambiguously and precisely. In this context, a system of language can be the official language of a country, the jargon for a professional community, or musical scores for musicians. From a communication perspective, it can be perceived as an overlapping realm of understanding that allows information to be successfully transmitted to all parties. Figure 18, for example, shows an intrapersonal version of Schramm's model (Fill 1999). A message was transmitted through a loop made by the two agents (illustrated by two circles). Due to their unique personalities, experiences, as well as environments, these two agents have different realms of understanding, but at the same time, they do have commonalities (illustrated by the overlapping areas of the two circles), such as holding similar values, living nearby, or

belonging to an identical scientific community, etc. Similarly, since both Bob and his fiancée live close to The Ohio State University, they have already developed an "overlapping realm of understanding" about the campus area. Therefore, whenever they encounter the term "Derby Hall", they picture the exact same place. By using this overlapping realm of understanding, Bob constructed a locational inconsistency between the geo-tag and his true location at the bar, whereas his fiancée constructed a locational consistency between the geo-tag and Derby Hall. Consequently, his fiancée was spoofed by the fake geo-tag.

In this sense, to understand location spoofing, it is imperative to determine whether an "overlapping realm of understanding" (or a shared code) exists in the course of transmitting the location spoofing, as well as its nature and function.

2.2.2 Ontology for understanding location spoofing: A critical realist perspective

To more deeply understand this "overlapping realm of understanding" as well as to investigate the hidden mechanism behind spoofing motivation, this dissertation will apply a more metaphysical framework to examine this spoofing phenomenon in addition to a linguistic one. Therefore, I will focus upon critical realism, mainly because it provides a realist ontology for interpreting scientific knowledge discovery (this dissertation is a typical discovering process for scientific knowledge about location spoofing). Critical

realism, initiated by Bhaskar (1975, 1979, 1993) and gradually improved by a few scholars (Archer et al. 2013), is a middle ground between positivism and interpretivism. For this reason, it is most suitable for investigating a research domain that encompasses both quantitative and qualitative practices. Sayer (1985, 1992) introduced critical realism to human geographic studies; Schuurman (2002), Perkins (2003), Leszczynski (2009), and Galt (2011) have examined GIS from a critical realist angle. Indeed, critical realism is commensurate with the growing needs of GIS criticism, rather than a mere positivist or empirical analysis of it.

Before applying critical realism to the phenomenon of location spoofing, I will begin with an introduction of this nuanced philosophical theory. In general, critical realism posits two fundamental ideas: (1) a world exists outside and independent of our conscious perception and (2) only some aspects of this world are objectively knowable via our senses. Bhaskar (1975) divided all knowledge in the world into two types: transitive and intransitive. The transitive objects include scientific concepts, laws, and theories. These objects are subjective as their existence relies on human activity; whereas intransitive objects are the "structures, mechanisms and processes, events and possibilities of the world; and for the most part, they are quite independent of us" (Bhaskar 1975, 22). Critical realism also proposes a three-layer ontology to further distinguish between what actually happens and what we perceive, and between an event and the underlying (but possibly unobservable)

mechanism. Figure 19 illustrates this stratified ontology as a nested relationship between three layers of the real, the actual and the empirical. The real domain is overarching and contains generative mechanisms and structures; the actual domain contains events that are instantiated by the mechanisms and structures; and the empirical domain primarily contains perceived phenomena, indicating those events that are observed or otherwise experienced. In this regard, mechanisms, events and experiences are real; events, experiences, and instantiations of the generative mechanisms are actual; experiences, which occur via empirical traces of actual events, are empirical.

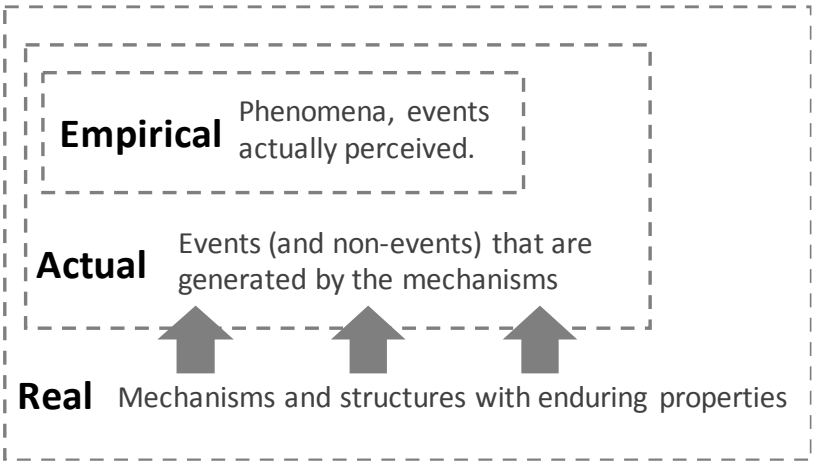

Figure 19. Stratified ontology of critical realism [Source:modified after Mingers (2004)]

More importantly, critical realism distinguishes between transitive and intransitive objects, especially their different roles in the discovery of scientific knowledge. While positivism/empiricism more generally reflects the causal relationship within the actual

domain, critical realism treats it as a generative mechanism, which belongs to the real domain. Critical realism deems that the causal relationship cannot be reduced to empirical constant conjunctions. The generative mechanism, the natural flow of things, is essentially the causal law. Unlike natural laws or regularities (belonging to intransitive objects), rules of society and human consciousness (belonging to transitive objects) are not universal but only applicable in a certain space-time, mainly since both human thoughts and society are continuously changing. Consequently, empirical constant conjunctions are unable to fully explain the generative mechanism that underlies the transitive objects, especially human motivations. So, the knowledge of an intransitive aspect can be quantitatively retrieved through empirical studies of its generative mechanism; whereas knowledge of transitive objects can be acquired via qualitative research, such as describing or interpreting its generative mechanism.

Therefore, location spoofing has been more richly conceptualized under a critical realist framework. By breaking down location spoofing into two separate facets, this study can avoid confusing the motivation with the inconsistency, and thus, establish appropriate epistemic frameworks to respectively manage these two aspects. First, the spoofing motivation is a typical transitive aspect related to human consciousness, the underlying generative mechanism of which can be interpreted rather than conjectured by empirical tests, natural laws, or trends. Hence, to uncover the spoofing motivation, I must analyze

location spoofing cases or interview location spoofers. In this way, I hope to more accurately describe the generative mechanism of spoofing motivation, because our current knowledge is lacking. Secondly, locational inconsistency can be regarded as an intransitive aspect of knowledge independent of human beings. In reality, this inconsistency can be objectively observed as a piece of locational information. Based on an "overlapping realm of understanding" in the GIScience and geography field, more broadly - the existing knowledge of geographic inconsistencies or impossibilities, an empirical study can be designed and implemented that will explain the disconnect between the observed locational information and the geographic ground truth.

2.2.3 Necessity of geographic knowledge for detecting location spoofing

Viewed from the GIScience perspective, location spoofing is apparently a serious spatial data quality issue. Accordingly, a review of the existing knowledge about spatial data quality assessments may shed light on the methodological framework for detecting location spoofing. Spatial data quality usually includes five elements: position accuracy, attribute accuracy, logical consistency, completeness, lineage and temporal quality; and these elements must be appreciated for their specific purpose and usage. When this definition is associated with location spoofing, the inherent locational inconsistency can be described by these five elements, and the spoofing motivation is closely related to the

purpose and usage. Hence, in this section I will review state-of-the-art volunteered geographic information (VGI) and geospatial big data quality assessments. I will stress the necessity of geographic knowledge for accurate detecting.

Today, we face increasing concerns about the popularity of location spoofing with the precise meaning shifting in data quality away from traditional concepts in the context of surveying and mapping to a more user-centric perspective. For example, emerging concepts like spatial data accuracy 2.0 (Goodchild 2008) and spatial metadata 2.0 (Kalantari et al. 2014) emphasize the fitness of purpose of locational information. Also, social and political credibility are becoming increasingly important for assessing locational information (Flanagin and Metzger 2008) rather than conventional technical quality and accuracy (Esmaili, Naseri and Esmaili 2013). In addition to the shift in the concept of spatial data quality per se, the increasing concerns about location spoofing also originate from the inherent suspicion of the accuracy of user-generated (geospatial) contents or namely VGI.

Multiple approaches have been proposed to assess the data quality of VGI. Goodchild and Li (2012) described three approaches, including crowd-sourcing, social, and geographic approaches. Specifically, the crowdsourcing approach, based on Linus's law - "given enough eyes, all bugs are shallow" (Raymond 2001), predicts that a crowd can collectively converge on the truth; a social hierarchy of trusted individuals maintains the data quality in

the social approach; last, the geographic approach is fundamentally based upon existing knowledge. According to Goodchild (2013, 284), these are the principles by which the geographic world is constructed. As an extension of these three approaches, Haklay et al. (2014) further explored three additional approaches to ensure VGI quality – the domain approach, the instrumental observation approach, and the process-oriented approach. Bordogna et al. (2014) studied a linguistic approach for VGI quality assessment. In addition, Loshin (2014) relied on maximum data usability to assess the quality of geospatial big data. From the perspective of cognition, all the approaches reviewed above can be considered as geographic, mainly because the very core of their criteria depends on a geographic interpretation about "how the geographic world is constructed". For example, both the social and crowdsourcing approaches provide a way to analyze and categorize individual experiences and judgments on a collective level. This method essentially relies on the assessor's own perceptions about what is geographically true or not.

Although elements of geographic knowledge are already in place to assess spatial data quality, we still lack a detailed theoretical framework to model these spoofing cases as well as robust methods for detecting these types of cases. Goodchild and Li (2012) have argued that a geographic detecting approach relies on a comparison of the observed locational information to the existing geographic knowledge. Some are known because they have been deduced from observations and through the scientific discovery process, while others

exist in the practices of regulatory agencies. Among the known portion of existing geographic data, time geography can be used as a foundation to formulate a method for detecting. The major reason is that time geography not only provides a series of human behavior constraints in space-time as its theoretical cognitive basis, but also develops efficient measurements to support concrete methods and practices. In the next section, I briefly introduce time geography to explain its association with spoofing detection.

2.3 Time geography

Inspired by conceptual advances in modern space-time physics (Kuklinski, 1987, pp.507), Hägerstrand (1970, 20-21) originated time geography in the mid-1960s with the purpose of "finding out the workings of large socio-environmental mechanisms" using "a physical approach involving the study of how events occur in a time-space framework". In his theory, each, individual has his or her own path in time-space, and life paths are hindered by several constraints. Though he and most time geographic researchers admit that "It would be impossible to offer a comprehensive taxonomy of constraints seen as time-space phenomena (Hägerstrand 1970, 10-11)", he has proposed three major types of constraints based on his own observations and empirical studies:

(1) Capability constraints are limitations on the activity of individuals due to their

physical/biological structure or available resources/vehicles.

(2) Coupling constraints define the spatiotemporal limitations of an individual, such as having to rely on other individuals or tools/materials to support producing, consuming and transacting.

(3) Authority constraints highlight the restricted space-time an individual is (in)capable of accessing. For example, access to a military zone is forbidden for average persons, and a grocery store only allows consumers to shop during working hours.

Overall, these three types of constraints encompass impacts of the natural environment, social norms, personal behavior and psychology, and provide fundamental hypotheses for time geographical measurements and analysis.

2.3.1 Binary models: Space-time prism

Conventionally, a binary concept in time geography defines an individual's mobility. Of all the binary models, the space-time (ST) prism shows individual's potential behavior in a time geographic context, which includes the capability and coupling constraints and highlights the influence of space-time anchors on the ability to participate in activities (see Figure 20). As shown, Anchors l and l' describe those fixed activities. The ST path, shown in red, represents the approximate trajectory of an individual through space and

time. The prism envelopes all possible space-time paths between two space-time anchors. The boundary of a prism is the object's maximum speed constraints imposed by the anchors. When the prism is mapped out in two-dimensional space, it forms a potential path area (PPA) (Miller 2005). The PPA is the projection of a prism constructed by two consecutive anchors, and is made up of up to two discs and one ellipse.

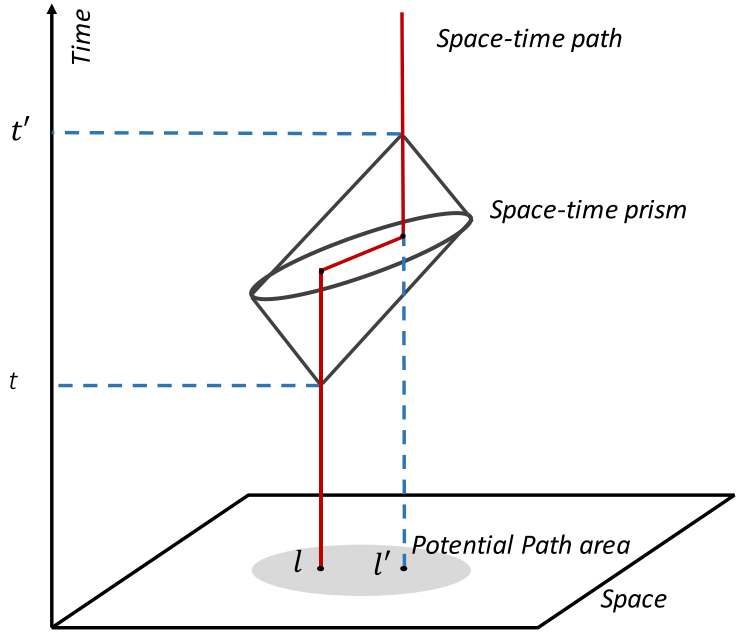

Figure 20. Time geographic measuring elements

Theoretically, it is impossible for an individual to be at a certain spot during a particular time budget if the path and the prism do not intersect. This strict limitation offers scholars an opportunity to measure the mobility of an individual based on a prism. Hence, since all possible paths are included within a prism and the inconsistent ones are left outside, the

prism provides an effective tool for sifting out location spoofing and other types of space-time inconsistencies.

2.3.2 Frequentist models: Prison/prism break

A prism, in a binary structure, implies that its interior is accessible and the exterior is not. Obviously, the texture of a prism is oversimplified and the boundary is actually mandatorily created. This over-simplification hides the uneven density inside a prism. In fact, a prism can be perceived as a three-dimensional structure, which will effectively stabilize various possible movements in space-time rather than a homogeneous object. Visiting probability indicates the probability at one specific location among all accessible locations. Winter and Yin (2010) have highlighted that the visiting probabilities inside the prism were not uniform. In other words, the distribution of visiting probability is unequal: regions near the axis are more likely to be visited than other regions near the boundaries.

In order to deal with the ball and chain brought about by the binary structure, several frequentist models, including, random walk (Winter and Yin 2010), Brownian bridges (Song and Miller 2014), and time-geographic density estimation (TGDE) (Downs 2010), have been developed to prison-break these inherent limitations.

First, Random walk is a basic visiting probability model based on the theory of random

walks. A random walk is a stochastic process in discrete space and time. At each time step, an individual walks in a randomly chosen direction, the visit probabilities of random walk follow a bivariate multinomial distribution centered on the prism anchor. However, movement is not completely random. So, a directed random walk, or a random walk process with directional biases, is created capture the most up-to-date movement behavior. A directed random walk turns out to be a visiting probability distribution similar to a bivariate normal distribution.

Second, the Brownian bridges theory estimates directed movement in continuous time and space. A Brownian bridge is a continuous stochastic process between two anchors, so it can capture directionally biased random movement between prism anchors. Since it is not constrained by a maximum speed and, therefore, is unaffected by the prism boundary, it is still possible for individuals to travel outside the boundary. A truncated Brownian bridges imposes a maximum speed on the Brownian bridges process and represents prism constraints (Song and Miller 2014).

Lastly, TGDE is another strategy for estimating visiting probabilities within a prism. It relies on time geography and density estimation to produce a continuous probability distribution of an individual's spatial position over time. TGDE assumes that the most likely locations to be visited are along the axis between the anchors. This method can generate detailed movement patterns of objects between anchors, which are derived from

sparse tracking data. This feature could be the most effective method for estimating the probability of activity of any individual social media users. Usually, the spatial prints from social media users are sparse in space and time. A social media user can either frequently post geo-tags or just occasionally send out a few, which can result in the distances between two consecutive locations varying from a few meters to thousands of kilometers, and also the time intervals in between varying from seconds to hours, and even to years. In addition, most prevailing location-based social media APIs provide a limited number of information entities, and consequently only sparse spatial prints can be collected for analysis. Considering the above issues, TGDE is best suited for formulating the visiting probability of social media users.

2.3.3 Bayesian models: A potential direction

TGDE, coupled with other frequentist models, requires the input of individual movement trajectories. Actually, apart from formulating individual movements, other macro controls can be used to improve the accuracy. Hence, to improve the current algorithm to holistically embody the three human constraints in space-time rather than simply individual activity, I propose a solution to improve the current estimation by employing Bayesian statistics. Such statistics, also referred to, as Bayesian theorem, are an aspect of probability theory that relates to conditional probabilities. A basic Bayesian method can be mathematically described in the following equation (Riesenfeld 2011):

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

For proposition A and evidence B, $P(A)$, the prior probability, is the initial degree of belief in A. $P(B)$, the prior probability of B, acts as a normalizing constant. $P(B|A)$ is the conditional probability of B given A. So, the quotient $P(B|A)/P(B)$ represents the support B provides for A, or namely the standardized likelihood. Therefore, $P(A|B)$, the conditional/posterior probability, is the degree of belief in A having accounted for B.

Bayesian statistics estimate parameters of an underlying distribution based on the observed distribution. It begins by modeling prior beliefs and updating them through evidences to obtain prior beliefs. In essence, Bayesian statistics provides a rule about how to update the strengths of prior beliefs in light of new evidence a posteriori. Placing time geography within the context of Bayesian statistics, the individual activity probability, representing individual capability constraints, can be regarded as the prior beliefs. If other human constraints can be integrated as new evidence, a Bayesian model can account for more human constraints in space-time, and link the prior belief with posteriori to determine the visiting probability of any individual.

Figure 21. All geo-tagged locations of Flickr photos [Source: Gilmer (2015)]

Conventionally, the coupling and authority constraints, derived from a top-down procedure, are modeled on the researcher's own perceptions about the world. In contrast, this dissertation adopts a data-driven approach that depends entirely on what the data per se represents rather than the perceptions about it. Simply put, these two constraints are predicted by the spatio-temporal dynamics of VGI rather than estimated from laws, regularities, or experiences. Indeed, the spatial distribution of an enormous amount of voluntarily derived geographic information largely represents how social media users are physically distributed on the Earth. For example, Figure 21 shows all the voluntarily derived geo-tagged locations of Flickr photos. Based on our common geographic knowledge and life experiences, we can affirm that places like the Sahara Desert or the Amazon Rainforest are less likely to be visited physically and be virtually geo-tagged in

location-based social media. In contrast, a data-driven approach disregards subjective perceptions as much as possible and relies on enormous geo-tags to estimate frequencies. Obviously, this data-driven approach can efficiently and effectively represent human constraints across the world at a collective level, and even unveil more hidden geographic facts in addition to the knowledge identified by experts. Therefore, combined with this data-driven approach, the Bayesian based detecting method would be well suited for sifting out location spoofing cases from normal locational information.

2.4 McLuhan's laws of media

Like all great minds before him, Marshall McLuhan's theory was ahead of its time. In retrospect, Marshall McLuhan was celebrated prematurely in the late 1960s, as his work is more relevant for the Internet than TV. His theory mainly focuses on how new technologies affect the thoughts and actions of individuals and society. A great number of scholars and critics have published numerous books and academic papers on McLuhan and his theories (O'Neill 2011, Levinson 1999, Theall 2001, Fishman 2006). In this sense, McLuhan's theory may be ideal for investigating the impact of location spoofing on society. The rest of this section gives a brief introduction to McLuhan's laws of media.

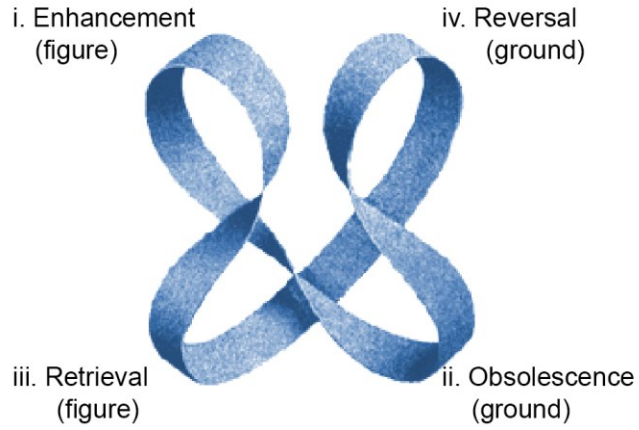

Figure 22. McLuhan's laws of media [Source: Sui and Goodchild (2003)]

According to McLuhan's media theory, human history can be divided into four phases, each of which has a dominant medium that produces a defining impact: the tribal age (oral and preliterate), the literary age (hand writing), the print age (printing), and the electronic age (electronic medium) (Griffin 2003). The complex impacts of media evolution are summarized as the laws of media and in the form of a tetrad (see Figure 22). McLuhan and Powers (1992) illustrated the four elements of a tetrad in a Mobius strip in order to show the simultaneous existence of these four elements. McLuhan's laws of media investigate how a certain medium or technology impacts human capacities and social practices in four ways: (1) enhancing, and simultaneously (2) making obsolete, (3) retrieving that which had been obsolete for a long time, and (4) reversing to the point at which the given medium or technology has been pushed to extremes.

To answer these four questions, McLuhan included two built-in ideas: the transition from

visual to acoustic thinking patterns and the figure-ground perception from Gestalt psychology. According to McLuhan and Powers (1992) , visual thinking (captured by the eyes and sensed by the left brain), reflects a linear and quantitative way of thinking, organizes information and has clear centers and margins, while acoustic thinking (dominated by the ears and processed by the right brain), reflects a holistic and qualitative way of thinking and is a dynamic process with "centers everywhere and margins nowhere". Drawing from figure-ground perception, McLuhan illustrated enhancement and retrieval as two figures and obsolescence and reversal as two grounds, thus showing the simultaneous, but not successive, hierarchical existence of these four elements. Therefore, by applying McLuhan's framework for studying location spoofing, I aim to develop a holistic view of the social impacts of this spoofing phenomenon.

2.5 Summary and research framework

This chapter reviewed the recent academic works related to location spoofing. According to cases reviewed, I define location spoofing as "an act of intentionally falsifying locational information to a place other than one's true location", and examined it as an information transmission process with several constituent elements (e.g., location spoofer, fake locational information, spoofed audience, etc.). Relying on a critical realist perspective, I emphasized the necessity of geographic knowledge for clearing up the

ambiguity in language fallacy and thus critically understanding location spoofing. Having laid a foundation for the importance of geographic knowledge, I further reviewed the relevant literature and research in time geography, especially the theoretical background, the binary and frequentist models, and finally suggested a potential direction by leveraging time geography via a Bayesian approach. Finally, to holistically understand the impact location spoofing may have on human society, I introduced the McLuhan's laws of media.

As a dissertation in the field of GIScience and geography more broadly, I would like to conceptualize my research under the general GIScience theoretical framework (Goodchild et al. 1999). Under this framework, three core areas of GIScience, making up three distinct areas of inquiry as an organic whole, aim to holistically understand what GIScience should address: (1) Cognitive models of geographical space: how we conceptualize the world around us and understand it through derived conceptualizations (2) Computational methods for representing geographical concepts: how we design a GIS solution to achieve maximum performance and utility, and simultaneously minimize information loss and/or overcome other constraints (3) Geographies of the information society: what mechanisms determine the adoption of GIS in society, and its use by human beings and how the adoption of GIS changes the way society perceives geographies.

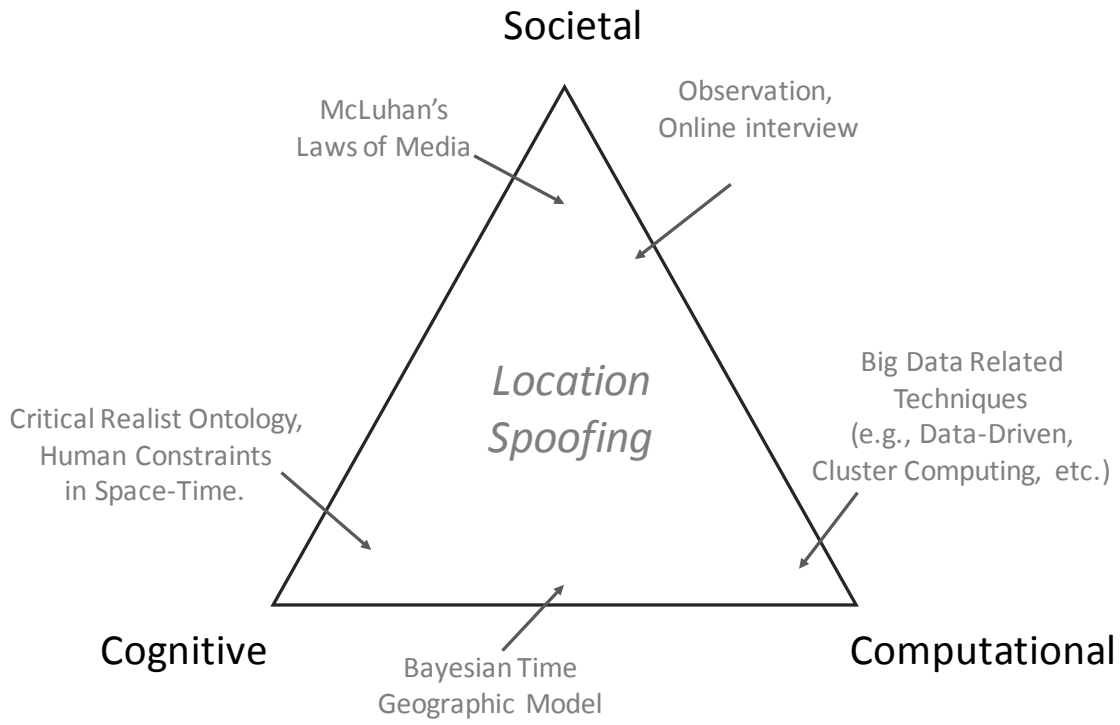

Figure 23. Research framework [Source: modified after Goodchild et al. (1999)]

Thus, by conceptualizing location spoofing under this GIScience theoretical framework, I have systematically developed a research framework for this dissertation, as shown in Figure 23. First, cognitively, this dissertation rests upon a critical realist ontology to determine what and how we can know about location spoofing. More specifically, it argues that location spoofing is essentially driven by motivations and manifests itself as a locational inconsistency between the geographic ground truth and the observed locational information. The necessity for geographic knowledge is seen as an "overlapping realm of understanding" to confirm what is geographically fake or not. Out of all existing geographic knowledge, this dissertation uses the three human geographic constraints

(proposed by Hägerstrand) as the cognitive basis for analyzing and interpreting this spoofing phenomenon. Secondly, computationally, by linking time geographic measurement (e.g., space-time prism, TGDE, etc.) to Bayesian statistics, a Bayesian time geographic approach is devised to separate location spoofing from geospatial big data. In addition, to manage a large-scale data set, I plan to integrate with relevant big data techniques, such as the data-driven approach for formulating human appearance probability and cluster based data harvesting information. Third, in reference to the societal aspect, I will unveil the generative mechanism of location spoofing – the nested motivation – through online observations and auxiliary interviews. Moreover, I will examine the social implications of location spoofing based on McLuhan's laws of media. Overall, the proposed GIScience research framework will guide me to holistically and critically examine this emerging spoofing phenomenon in social media as well as other forms of geospatial big data.

Chapter 3: Methodology

In this chapter, I propose a hybrid methodology that embraces both a quantitative approach to detect location spoofing in social media as well as a qualitative investigation on the hidden motivations for this activity. The detecting approach utilizes time geography to model the spatial-temporal constraints of human activities and to sift out locational inconsistencies and impossibilities. This human constraint is well accepted: an individual, regardless of the selected mode of travel (e.g., walking, biking, driving, taking the train or flying, etc.), is constrained by the speed limit of available transportation vehicles. So, during the time it takes to post two locational information entities, if the average speed is beyond the limit (e.g., the maximum speed limit of the fastest civilian airplanes, or the local ground vehicle speed limits under specific circumstances), I can confidently infer that at least one of the two entities is spoofed. Subsequently, I can calculate the individual's visiting probabilistic value at the two locations of interest, respectively. As a result, the one with a lower value is more likely to be a fake. Secondly, through in-depth online observation and auxiliary interviews, I am able to utilize the qualitative approach to create elaborate profiles of the possible location spoofers and collect evidence related to the generative mechanism (causation) that may trigger the spoofing phenomenon. Finally I

deeply investigate, the motivation for spoofing and some of its more controversial aspects.

Before elaborating upon the procedure for the proposed method, I had to overcome difficulties of acquiring, parsing and storing geospatial big data. Admittedly, nearly all big data studies must overcome these difficulties. Thus, the chapter begins with an introduction on how to crawl large-scale social media feeds and then moves to an exploration of the detecting algorithm based on the collected data. In addition, this chapter also summarizes the investigating procedure for profiling the spoofing motivations.

3.1 Crawling and harvesting social media feeds

Like most research focusing on social media data, an efficient data crawler ensures that the study is scientifically reliable. A data crawler (also known as a web spider or web robot) is an automated script/program that methodically gathers data from web resources. Usually, harvesting social media feeds via a web crawler consists of three steps: (1) Retrieving data from social media providers; (2) Parsing and storing the data; and (3) Analyzing the stored data (Croitoru et al. 2013). There are two primary approaches to retrieving data from social media. One is to obtain the visited HTML pages and then extract the structured data from the page. Another is to directly retrieve the data in a structured format (e.g., xml or json) through APIs (Application Programming Interfaces). Compared to HTML requests, APIs

are quicker to respond to data; therefore, the crawled results are much more efficiently structured.

I used Twitter feeds as an example to help me explain the procedure of data crawling, mainly because Twitter is a prevailing social media resource with users all over the world, straightforward APIs, and plenty of geo-tagged "tweets". In this context, a tweet refers to a short text message with no more than 140 characters. Each tweet includes data fields, including content, metadata (id, retweet count, etc.), the user's profile (username, id, screen name and etc.), language, source (the client or platform by which to send a tweet), the created-at timestamp, latitude and longitude. Launched in July of 2006, Twitter rapidly gained worldwide popularity, handling as many as 1.6 billion search queries per day (Twitter 2011). As of December 2014, Twitter had more than 500 million users, 284 million of which are active users (Keach 2014). Users can access Twitter through the website interface or mobile device app, In addition, Twitter provides APIs to enable third-party web applications to post tweets from their customized client lists and link to the Twitter database. In this way, social media users can send out geo-tagged tweets from multiple sources/platforms, such as Instagram, Foursquare, Flickr, Path, etc.

To test the detecting algorithm in a common big data setting, I configured the data crawler to complete the following steps: (1) Due to the fact that a large data set of numerous geo-tagged tweets is required for screening, I relied on the Twitter's public stream API for

a representative set. By adding a bounding box parameter to the URL request, I ensured that this API can continuously harvest geo-tagged tweets. For example, the list value [-180, -90, 180, 90] allows the crawling program to efficiently retrieve geo-tagged tweets from all over the world. Then, I stored all the harvested geo-tagged tweets into a spatial database (e.g., SpatiaLite, PostgreSQL with PostGIS, Oracle with Spatial extension, etc.). (2) Subsequently, for each geo-tagged tweet in the spatial database, I was able to acquire a consecutive geo-tagged tweet from the same user by Twitter API's user timeline function. Having access to the tweet id and the user id (the user sends out the geo-tagged tweet), I obtained up to a maximum of 200 most current tweets for each request[9]. The set of retrieved tweets satisfies two conditions: they were all from the given user and were posted before the given geo-tagged tweet. Then, I filtered out the geo-tagged ones and specified the one closest in time as the consecutive geo-tagged tweet. (3) Later, I separated and stored the geo-tagged ones from the 200 tweets to determine the user's activity in space and time.

The first step (acquiring a representative data set as the candidates) only required one request, and then the data would be fed back in a timely manner. However, for the last two steps, the crawler must send a request for each geo-tagged tweet. The staggering numbers of geo-tagged tweets in the data set frequently overload the server. Therefore, most web

[9] Refer to https://dev.twitter.com/rest/reference/get/statuses/user_timeline.

applications limit requests for excessive data from one computer. Twitter, for example, has a very strict API rate limit rule[10]:

> "While in version one of the API, an OAuth-enabled application could initiate 350 GET-based requests per hour per access token, API v1.1's rate limiting model allows for a wider range of requests through per-method request limits. There are two initial buckets available for GET requests: 15 calls every 15 minutes and 180 calls every 15 minutes."

With these limits in mind, Twitter allows no more than 180 calls per 15 minutes. If so, the second step may require roughly a month to process a million geo-tagged tweets. To make the crawling procedure more efficient, a data crawler can divide a complex problem into smaller portions and process them separately. In fact, this is one of the core advances for manipulating big data, which follows a philosophical principle known as "MapReduce". By definition, MapReduce "is a programming model for processing and generating large data sets with a parallel, distributed algorithm on a cluster" (Dean and Ghemawat 2008, 107). A cluster allows the master node to assign separate harvesting tasks to each cluster slave, and then each slave harvests a specific portion of data. When each slaves completes its task, the cluster assembles all harvested portions into one well-organized data set.

[10] Refer to https://dev.twitter.com/rest/public/rate-limiting

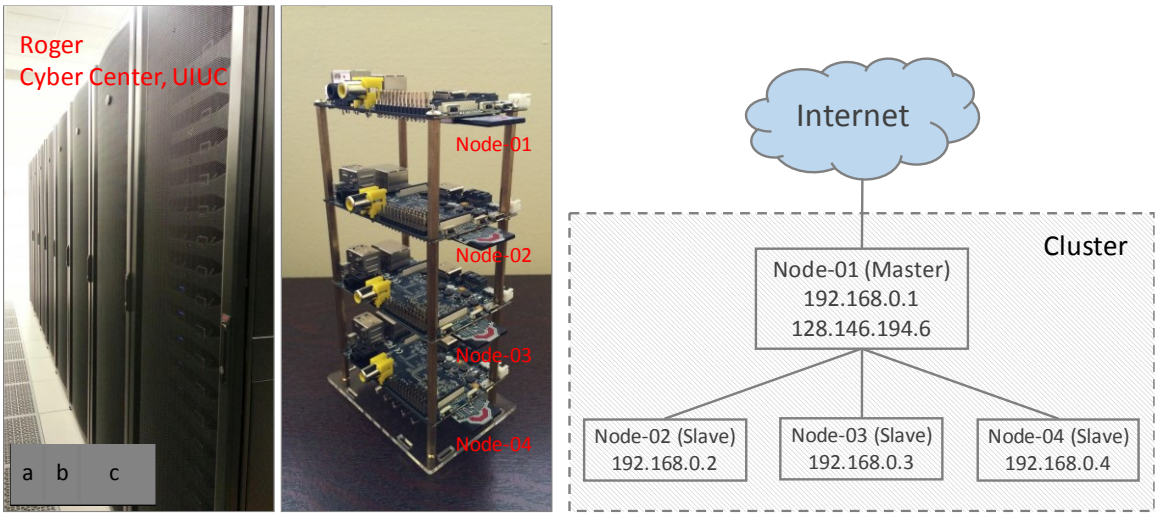

Figure 24. Open-source data harvesting cluster

CyberGIS scholars are dedicated to assembling a high-performance cluster to handle the massive amounts of geospatial data (GCN, 2015). As shown in Figure 24 (a), ROGER (Resourcing Open Geospatial Education and Research) cluster, named after Roger Tomlinson, is a device funded by an NSF grant (Award # 1429699) of $1.5 million over three years (GCN 2015). ROGER will be equipped with more than seven petabytes of raw disk storage, solid-state drives, and advanced graphics processing units, a high-speed network and dynamically provisioned cloud computing resources. Since ROGER remains an ongoing CyberGIS project, at this point, it cannot yet accept public usage applications from other organizations. Therefore, I built a cluster for processing geospatial data from scratch. The designed cluster, as shown in Figure 24 (c), consists of four computers (nodes), a router, four cables, a UPS and a patch board. According to the prices on Amazon.com in early 2015, it cost approximately one thousand dollars. If I chose some high-quality parts,

the price would double. Compared to x86 computers, the open source arm7 chip is less expensive but has equivalent functionality (e.g., Linux compatibility). In order to save money and make the cluster accessible for common academic use, I decided to choose banana Pi, a typical arm7 chip, as the node. A banana Pi cost no more than $40[11], and a cluster, as shown in Figure 24 (b) is no more than $300 in total. In addition to its affordable price for academic or personal use, this cluster is much easier to alter to the needs of the user. In other words, by simply adding or removing nodes, this cluster is able to harvest big data on different scales. Therefore, this cluster will allow me to efficiently harvest large-scale samples of locational information from social media.

3.2 Detecting fake locational information

To elaborate on the detecting procedure, an arbitrary locational information entity from location-based social media is denoted in the equation below:

$$s = (l, t, u, o) \quad (1)$$

Where s indicates an arbitrary information entity, u denotes the social media user that sends the information entity s out, l is the locational components of s, indicating the place in the real world where s has been created, while t is the timestamp, showing when

[11] Refer to http://www.amazon.com/Banana-Mini-Open-Source-Mainboard/dp/B00MWA5JKM

s was created, o indicates any other attributes of s.

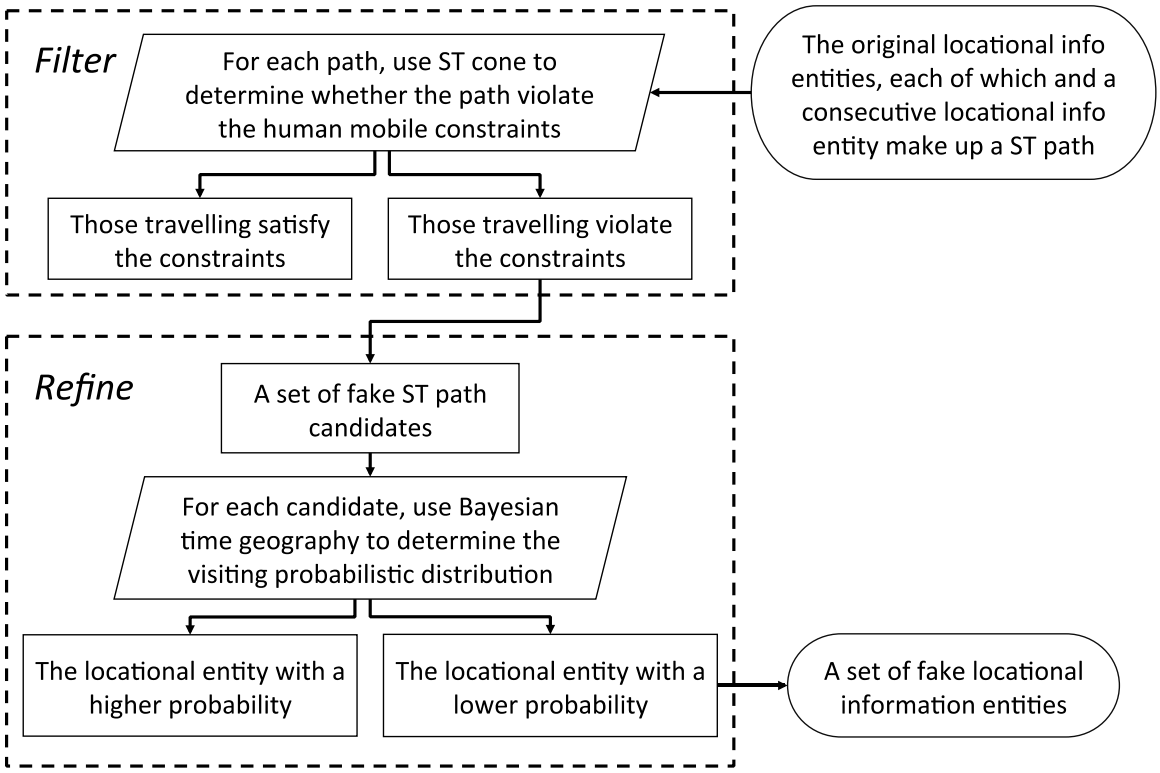
Filter
For each path, use ST cone to determine whether the path violate the human mobile constraints
The original locational info entities, each of which and a consecutive locational info entity make up a ST path
Those travelling satisfy the constraints
Those travelling violate the constraints
Refine
A set of fake ST path candidates
For each candidate, use Bayesian time geography to determine the visiting probabilistic distribution
The locational entity with a higher probability
The locational entity with a lower probability
A set of fake locational information entities

Figure 25. Flow chart for detecting fake locational inforamtion

The algorithm to detect fake locational information l_s is based on the filter-refine strategy (Wood, 2008). In general, this strategy consists of two steps. First, a set of candidates is filtered from the original data. This new set of candidates is a superset of the answer set, which consists of both the actual answer and the false ones; Secondly, each candidate from the filter step is then analyzed according to some other constraints in order to produce the answer set by eliminating the false ones. This strategy is employed as shown in Figure 25. Firstly, The detecting algorithm singles out a set of ST path candidates that violate human

activity constraints through ST cones. Then, for each ST path candidate, a Bayesian time geographic approach is conducted to calculate the probabilistic distribution of visiting different places in the research area for the author of the path. After comparing the probabilistic values of the two pieces of locational information (which link to the ST path), I deem the one with a lower visiting probability is more likely to be the fake locational information. In the following paragraphs, the flowchart is explained in greater detail.

According to Hägerstrand (1982, 331), […] an individual is restricted by the location in time and space of fixed points which must be respected. The time-spaces left free are defined by more or less symmetrical double cones, called prisms." Whilst explaining the prism concept, Hägerstrand devised a space-time (ST) cone. Hence, as shown in Figure 26, affixed to one anchor l, a cone extends itself in space-time, its interior portion features the accessible portion. The higher the maximum speed V_{max} of an individual u, the larger the interior portion (or PPA from a two-dimensional perspective). Thus, if the other anchor l' is outside the cone, I can then assert that l' is inaccessible to u, or in other words, the given space-time path p linking l and l' is a fake one p_s. Notably, this mobility constraint by no means can detect all types of location spoofing. For example, it is less effective when continuous pieces of locational information are sent out at a fixed spot, or if a fake path is plotted with a reasonable speed, the geographic constraint is unable to detect it. Despite the above limitations, the time geographic framework automates the detecting process, and a

number of location spoofing cases can be separated from large-scale social media feeds.

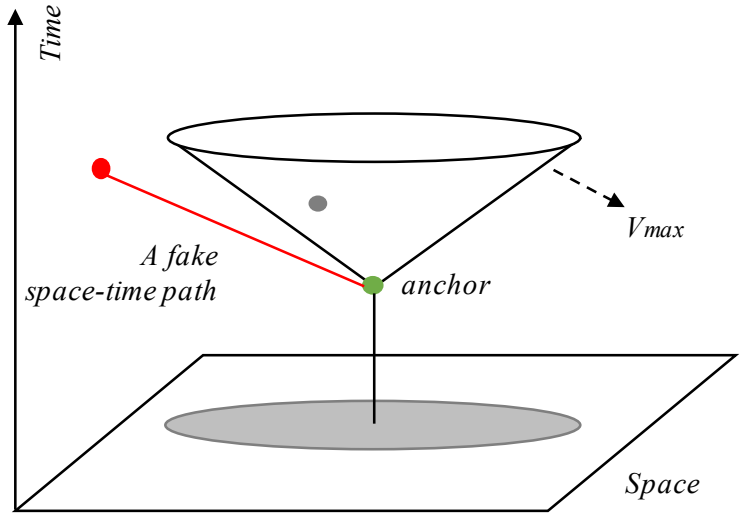

Figure 26. Detecting a fake path via a space-time cone

In order to detect the fake one from the two pieces of suspicious locational information, I plan to compare their visiting probabilities. The one with a lower value is more likely to be fake For a given individual social media user u, $P(V)$ is the visiting probability at an arbitrary location l, which can be formulated as the equation below:

$$P(V) = P(A|H) = P(A) \cdot \frac{P(H|A)}{P(H)} \quad (2)$$

Where $P(A)$ is the prior beliefs, indicating the individual activity probability. $P(H)$ is the posteriori probability, indicating the probability of human appearance, which is modeled

mainly based on authority and coupling constraints[12]. Here, $P(H|A)$ is defined as a piecewise function:

$$P(H|A) = \begin{cases} P(H) & D > p \\ P(A) & 0 < D \leq p \end{cases} \quad (3)$$

Where D indicates the density of locational information entities sent out at the measuring location l during a specific period, which can be days, weeks, months or even years. p is a threshold to differentiate crowded places from sparse (or deserted, desolate, hash, uninhabited) places. When $D > p$, meaning l is a crowded place, human movements at a collective level would be barely affected by the activity of an individual. At this point, $P(H)$ and $P(A)$ are two independent events, therefore $P(H|A) = P(H)$; When $0 < D \leq p$, meaning l is a sparse place, human movements at a collective level in such a place would be significantly affected by the activity of an arbitrary individual. So, I assume that $P(H|A) = P(A)$ at this point. Ideally, D can be interpolated by a kernel density function. Alternatively, to decrease the computational complexity, an efficient method is to count the total number of locational information entities within a fixed size window centered at the measuring location l. Based on my observation, the distribution of D forms a long-tailed curve. In the end, $P(V)$ can be formulated as:

[12] The connotations of *P(A)* and *P(H)* are discussed respectively in Section 2.3.3

$$P(V) = \begin{cases} P(A) & D > p \\ \dfrac{P(A)^2}{P(H)} & 0 < D \leq p \end{cases} \tag{4}$$

Having determined the visiting probability, I continued to formulate the individual activity probability $P(A)$ and the human appearance probability $P(H)$ in details.

3.2.1 Prior beliefs: Individual activity

The individual activity probability $P(A)$ can be formluated using TGDE, which enables us to display movement patterns of human travel in two-dimensional Euclidean space. This is particularly helpful for cases in virtual space, such as this research. By fitting a distance-weighting function to each geo-ellipse, TGDE creates a smooth density surface that measures the probability that an individual will travel to an arbitrary location l during a specific time. It can be formulated as:

$$P_l(A) = \frac{1}{(N-1)[t(1,\ N) \cdot v(u)]^2} \sum_{i=1}^{N-1} G\left(\frac{d(i,l) + d(l,i+1)}{t(i,i+1) \cdot v(u)}\right) \tag{5}$$

Where $P_l(A)$ is the probability that an individual will perform a certain activity at an arbitrary location l in the research area, N is the number of harvested locational information entities sent by the individual, where consecutive points are denoted as i and $i+1$, t is the time frame between sending out the two locational information entities i

and $i+1$., v is the maximum transmitting speed of an object, G is a distance-weighting function of the PPA, and d is the distance between location l and another anchor.

To use this formula for calculating the activity probability of an individual social media user, a set of consecutive locational information entities l from the given user u are retrieved. The total number is N. To make sure the closer places have a higher likelihood of activity than those farther away, I specified G as an inverse distance weighting function, and then checked that the probabilistic values equal to zero outside the PPA. As indicated by Fotheringham (1981), if $G(d)$ is used to predict human activities, r is between 1 to 2:

$$G(d) = \frac{1}{d^r} \tag{6}$$

The weight of the function of each geo-ellipse is acquired by computing the minimum distance necessary for the given user u to travel from anchor i to an arbitrary location l and then to anchor $i+1$, and to place this value over the maximum distance this user u could have possibly travelled during the time frame between anchor i and $i+1$. In order to obtain the maximum speed v, I calculated the speed of transmission between each pair of two consecutive locational information entities, and determined the highest value as the maximum speed $v(u)$. The remainder of the formula is divided by $(N-1)$ and the square of the maximum travel distance $[t(1,\ N) \cdot v(u)]^2$.

3.2.2 Posteriori probability: Human appearance

The human appearance probability $P(H)$, indicating how often humans collectively travel to an arbitrary location, can be calculated with minor modifications through TGDE, which is specifically formulated as:

$$P_l(H) = \frac{1}{M[t(1,\ N) \cdot v(u)]^2} \sum_{j=1}^{\mathrm{M}} G\left(\frac{d(j,l)}{rt(j,l) \cdot v(u)}\right) \tag{7}$$

Where $P_l(H)$ is the human appearance probability at any arbitrary location l within the research area, M is the number of existing locational information entities. $v(u)$ is the maximum possible speed of an object, G is a distance-weighting function of the PPA, and d is the distance between location l and another locational information entity j . $rt(j,l)$ indicates the time frame between location j and location l, which is determined by finding the temporal resolution under investigation. For instance, if the resolution is set as the time of day, the time frame will be calculated based on the time differences between j and l. If the resolution is set as a weekday, I need to also consider the number of days in question (i.e., the difference between Monday and Wednesday is two days).

Similar to $P(A)$, t and d are calculated the same way, and N and $v(u)$ are specified as the same values used in $P(A)$. But unlike $P_l(A)$ requires a set of locational entities that are continuous, these entities for $P_l(H)$ are from multiple users and may not be continuous. Its total number is M. Regarding the distance-weighting function G, the variable for each

geo-ellipse is acquired by computing the minimum distance necessary for the given social media user to travel from anchor j to an arbitrary location l, and place this value over the maximum distance the given social media user u could have possibly travelled during the time frame between anchor j and l. r is specified as the same value used in equation (6). Having figured out the fine-details, $P_l(V)$ can be formulated as:

$$P_l(V) = \begin{cases} \dfrac{1}{(N-1)[t(1,\ N)\cdot v(u)]^2}\displaystyle\sum_{i=1}^{N-1} G\left(\frac{d(i,l)+d(l,i+1)}{t(i,j)\cdot v(u)}\right) & D > p \\ \dfrac{M}{(N-1)^2[t(1,\ N)\cdot v(u)]^2}\dfrac{\sum_{i=1}^{N-1} G\left(\frac{d(i,l)+d(l,i+1)}{t(i,j)\cdot v(u)}\right)}{\sum_{j=1}^{\mathrm{M}} G\left(\frac{d(j,l)}{rt(j,l)\cdot v(u)}\right)} & 0 < D \le p \end{cases} \tag{8}$$

Therefore, for a fake path p_s that has starting and ending anchors which are two pieces of suspected fake locational information, I assumed the one with a lower visiting probability is a piece of fake locational information l_s. Of course, the detecting approach would definitely tolerate outliers. Based on the preliminary tests, it is possible to obtain no locational information entities with which to compute the individual activity probability. In this case, the detecting approach would be simply to compare the human appearance probabilistic values at those two places; as stated above, the one with a lower value is more likely to be falsified. Moreover, if the visiting probability values of the two locational entities are equal, both of them are assumed to be fake ones.

Admittedly, this detecting approach has some limitations. For fake locational information

that is being continuously sent out at a fixed spot or does not violate human mobile constraints, this approach is not applicable. Furthermore, it was not my goal to develop a versatile approach for detecting any location spoofing in social media - an extremely challenging (if not entirely impossible) task.

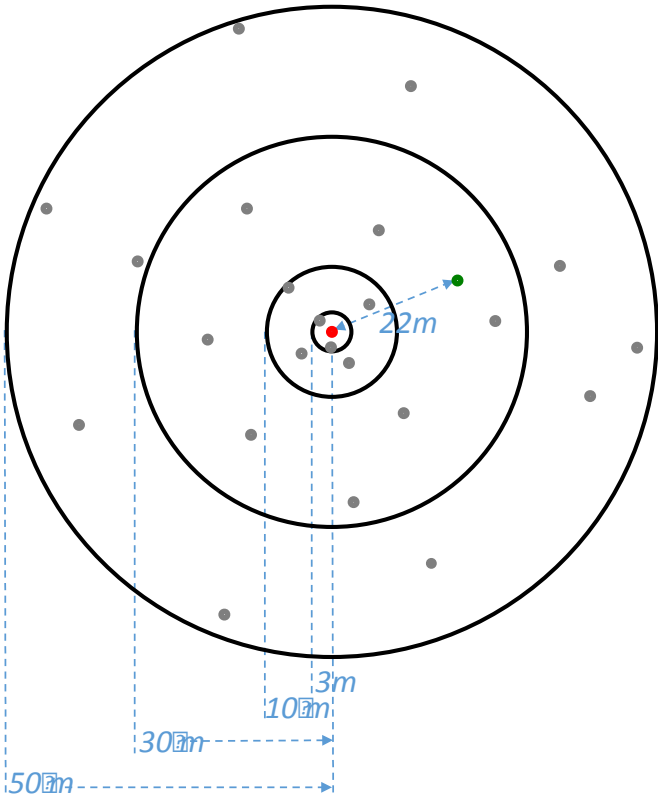

Figure 27. Measuring errors from a mobile device

Moreover, even if an ST path violates the human mobility constraints, under some circumstances, we cannot simply assume that it is fake. Possibly, the deviation between the two pieces of consecutive locational information entities is caused by the measuring errors.

To be specific, since the mobile device relies on hybrid positioning measurements, the measuring accuracy ranges from three meters to 50 meters (more details in Section 4.4.2) depending on the specific positioning measurement being used. In other words, even if two pieces of consecutive locational information entities are sent out from the same place via the same mobile device, there is still a possibility that the information could be offset in between. As shown in Figure 27, even if the user is stranded at the same location, the mobile device may capture two different geo-tags (red dot and green dot) that are 22 meters apart (The gray dots are other possible locations). Undeniably, this measuring error is a type of fake locational information, but there is no explicit deceptive intention associated with it. Therefore, I cannot refer to it as location spoofing. Confronted with a set of detected fake locational information, I need to understand the hidden motivation to determine which fake locational information is spoofed.

3.3 Profiling spoofing motivation

Rather than relying on a quantitative approach, this dissertation uses an in-depth qualitative investigation to study spoofing motivation and other possible generative mechanisms (causation). This investigation begins with observing the person suspected of location spoofer and then interviewing him or her online if necessary. Having collected sufficient information, this investigation can elaborately profile the suspects, and then critically

analyze their spoofing motivations as well as find out other possible generative mechanisms in the spoofing process.

Specifically, using each piece of detected fake locational information, I profiled the suspected location spoofer through a detailed online observation. Throughout the course of observation, I carefully looked for any useful materials related to the fake locational information as well as any evidence which might shed light on its author's motivations. Under most circumstances, this observation is conducted through browsing the author's Twitter web page, the author's record in the database, and any other Internet resources. In the end, the observed evidences are organized into an information portfolio. Generally, this information portfolio contains four types of information (ordered by the size of the information) in terms of auxiliary content, the author, the author's virtual community, and the broader context. Take a fake geo-tagged tweet for example, (1) the auxiliary content would be all the data fields of a geo-tagged tweet except the attached geo-tag, including textual content, timestamp, source, language used and so on; (2) the observed information related to its author includes its profile (e.g., username, textual location, joined date, personal web site, etc.), all (geo-tagged) tweets, and favorite posts; (3) the virtual community indicates the social network formed by the author's followers; and (4) the broader context indicates other relevant information from Twitter and/or other sites. Having collected these various types of structural data as an information pool, it is

necessary to dive in and carefully search for any possible clues that might explain the generative mechanisms of the spoofing phenomenon. However, it is worth noting that a complete profile is unnecessary and impractical. Under most circumstances, a geo-tagged tweet per se or a combination of several pieces of information may be enough. If all the clues still cannot reveal the motivation, I may conduct an online interview with the suspected location spoofer through Twitter Direct Messages (a built-in instant chatting service). During the interview, I would directly ask the interviewee about his or her motivation, which would enable me to efficiently identify the hidden motivation. Overall, a detailed profile with sufficient evidence may reveal motivation.

Moreover, if this process cannot firmly identify a user as a location spoofer, it is reasonable to investigate other potential suspects. Actually, locational information can be falsified during the entire course of information transmission, rather than simply at the critical final step of pressing the "send" button. Other stakeholders, such as social media providers and market regulators (government/militant agencies) also participate in spreading fake locational information, and consequently their motivations are closely intertwined with one another. Therefore, all these stakeholders and the author, shall be recognized as "conspirators" rather than direct spoofing masters (location spoofers). In this sense, all the conspirators are investigated through profiling as well.

3.4 Summary

This chapter explored a hybrid methodology for studying location spoofing. Based on large-scale social media feeds harvested by a specifically-designed open source cluster, a quantitative approach was proposed to detect location spoofing utilizing a Bayesian time geographic approach. Later, a quanlitative approach was introduced to determine spoofing motivations through online observation as well as aulixiliry interviews. To sum up, the detecting procedure is illustrated in Figure 28 and exemplified with a real case in Twitter.

In Figure 28 (a), the red dot, indicating Midtown Manhattan in New York city, is an arbitrary geo-tagged tweet retrieved from the public stream API. This tweet was sent out at 8:4:22 PM EST, on February 10, 2015. Based on my detecting procedure, I, then, must acquire another geo-tagged tweet that was sent out by the same user in the same time frame. These consecutive tweets (the green dot), originating from the Citi bike station, were sent merely six seconds before and approximately one kilometer away at the shortest distance. If the user were capable of travelling between these two places and then sending out the tweets, he or she must be travelling at the super human speed of 748.85 km/h, apparently, way beyond human physical capability. Therefore, I can confidently affirm that it is a typical spoofing case, where at least one of the two geo-tagged are falsified. Subsequently, to determine the fake one(s), I must compare the probabilities of this user visiting these two places. To do so, I queried 200 sequential tweets satisfying two conditions that (1) the

tweets are from the same user who posted the two suspicious geo-tags, and (2) the tweets are closely previous to the two suspicious. Since the user is an enthusiastic fan of location-based services, the 200 tweets are all geo-tagged.

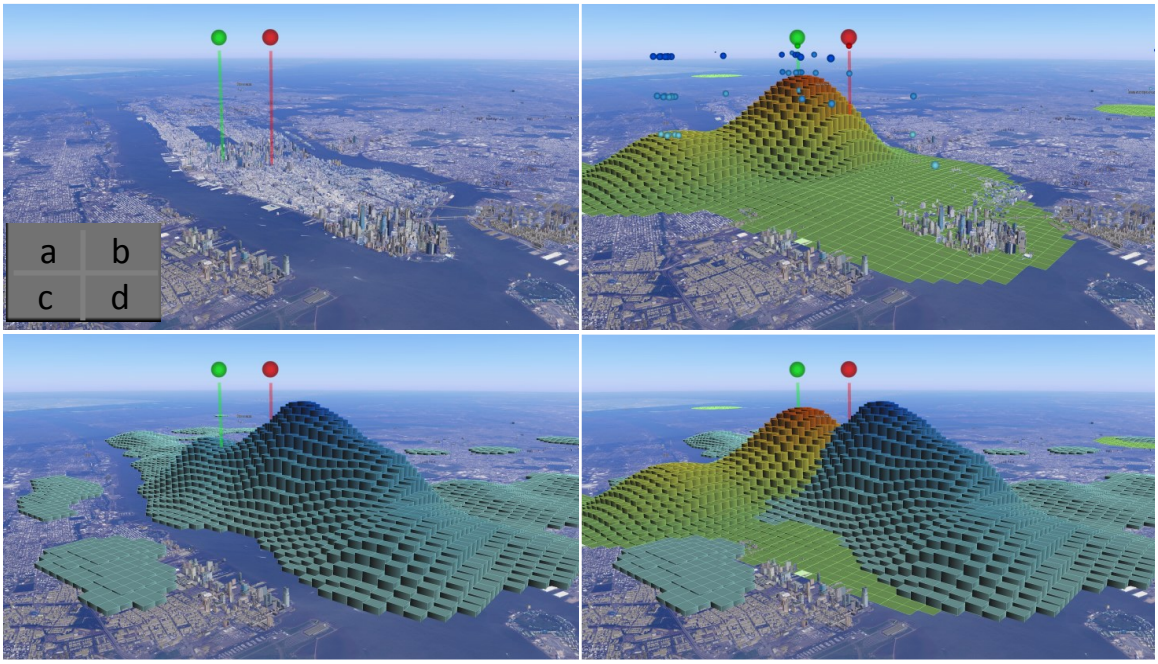

Figure 28. A real location spoofing case in New York city

Consequently, I can compute the spatial distribution of individual activity probability, thereby visualizing it on the ground as a 3-D mountain-like probabilistic distribution in Figure 28(b). Apparently, this user is more active in the place with a high elevation than low. Therefore, this user is more active at the green dot than the red. Additionally, with all the geo-tagged tweets around the two suspicious locations in Manhattan region, the human appearance probability can be computed as well. As shown in Figure 28(c), the human

appearance probabilistic values for are slight different with each other. But, both in the most crowded area in New York city, the two locations are all in the crowd places, meaning, the human appearance would not conspicuously affect the travel of the given user, thereby the visiting probability at these two locations will be determined mainly by the activity probability of this user. Hence, since the red dot has a lower probability of activity, I can confirm that the geo-tagged tweet at Midtown Manhattan (the red dot) is more likely to be a fake one. After singling out this fake geo-tagged tweet, I continue to investigate the motivation for spoofing. To do that, I need to profile the author of this fake geo-tag through all possible means, such as obtaining all data fields of this tweet from the database, browsing his or her Twitter web page, or discovering this user's Twitter followings and followed users. In addition to the above observations, I may also need to interview this user if my observations cannot fully explain the spoofing motivation. With this online observation and auxiliary interview, the motivation for spoofing can be elaborately investigated.

Chapter 4: Empirical Results

In this chapter, I implement an empirical study to examine location spoofing via a hybrid method. The primary goal of this empirical study is to test the effectiveness of the detecting approach utilizing millions of geo-tagged tweets and to reveal the hidden motivations for spoofing through online observation and interviews. Specifically, this chapter begins with an exploration of the data set followed by a description of a test of the procedure using this data set. Having successfully singled out a set of fake locational information, I examine the statistical, spatial-temporal characteristics of this spoofing phenomenon and evaluate this detecting approach to determine its limitations and possible direction for further improvement. Moreover, I investigate various hidden motivations for spoofing and any other generative mechanisms which may lead to the spoofing phenomenon. Finally, as an emerging topic in the era of big data, location spoofing has largely been ignored by society, the consequences of which can be profound (and in some cases deadly). To develop a holistic understanding of location spoofing in social media, I plan to place this study within the broader framework of McLuhan's laws of media.

4.1 Settings: Data and variables

Continuously running for roughly two days and five hours from 2015-02-09 00:09:35 to 2015-02-11 05:04:28 GMT (Greenwich Mean Time), the crawling cluster harvested 2,379,980 pairs of geo-tagged tweets, which were stored into a spatially enabled database (e.g., SpatiaLite, PostgreSQL with PostGIS, Oracle with Spatial extension, etc.).

4.1.1 Description of the data

By mapping the harvested geo-tags to a particular geographic space (see Figure 29), I could follow the movements of the Twitter users. Though Twitter alleges to be a global online community, the spatial distribution of geo-tweets that reflects the online activities on Twitter is extremely uneven (Graham and Zook 2011). Understandably, both natural environments and social-economic development level may influence the geographical distribution of the cyber-infrastructures and consequently expand the boundaries of the geo-tagged tweets. As a result, geo-tags seldom appear in sparsely populated areas, such as near the oceans, the Poles, the deserts or the forests. Moreover, in several authoritarian counties such as China, Turkey, and North Korea, Twitter is regarded as a huge threat and is blocked by the governments nationwide. Even through the citizens of these countries (e.g., China) have access to high-speed Internet, I have located only a few geo-tags from these areas. At the global level, few users are in East Asia (except Japan and South Korea),

the Middle Asia and most of Africa, while tweeting is most common in America and Europe.

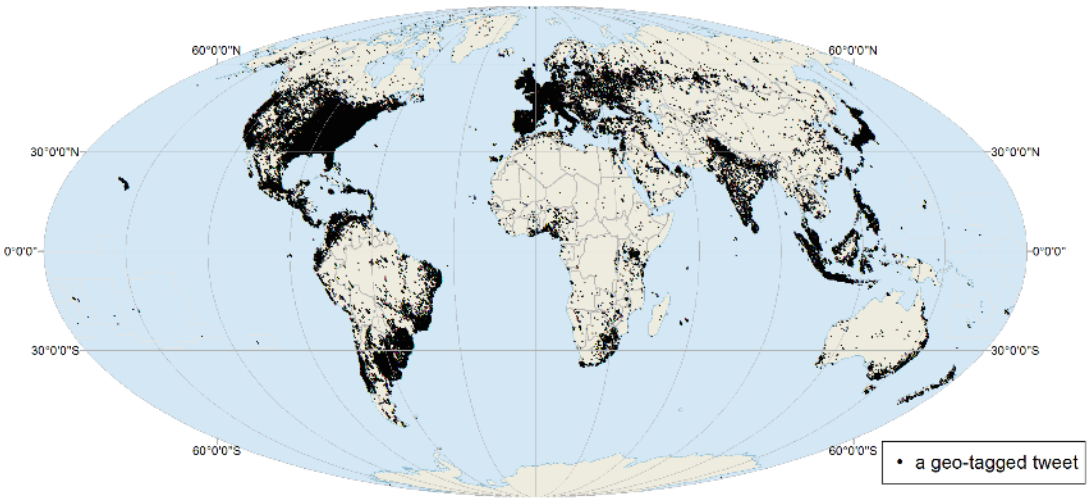

Figure 29. Global distribution of all harvested geo-tagged tweets

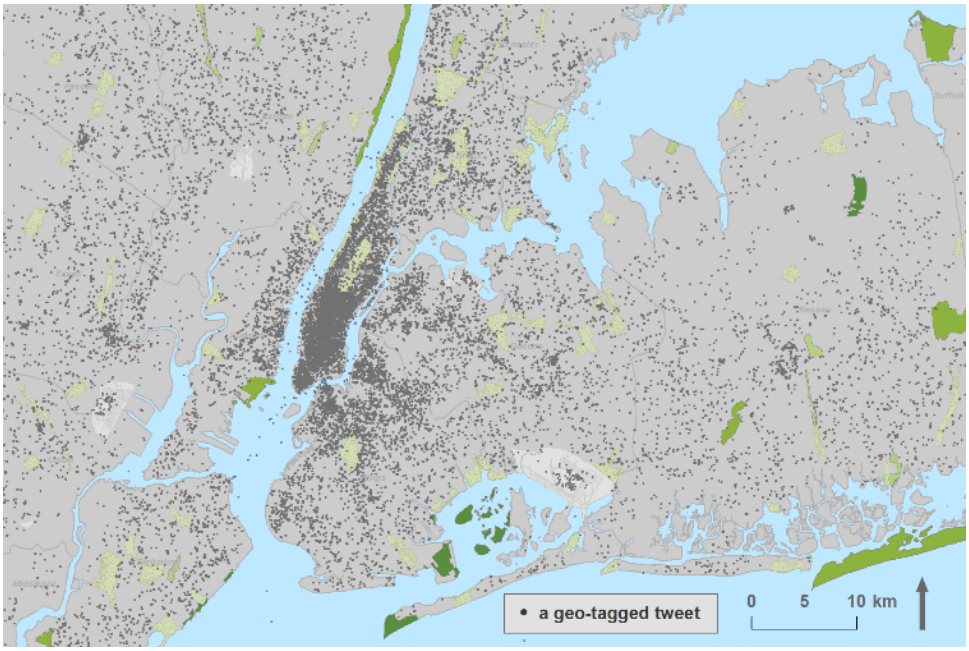

Figure 30. Harvested geo-tagged tweets around NYC

When we focused on a smaller range, we found that natural environments and human activities also restrict the spatial distribution of geo-tagged tweets. In New York City (Figure 30), for example, most of the geo-tagged tweets are within Manhattan except the central park. As a recreation place for New Yorkers, the central park only attracts fewer geo-tags. Likewise, there were only a few geo-tags to be found on the rivers (e.g., Hudson River, East River, Lower Bay), in the parks (e.g., Richard DeKorte Park) or other recreation areas (e.g., Gateway National Recreation Area). In addition, geo-tagged tweets were concentrated at major public service locations, such as airports like Newark Liberty International Airport, John F. Kennedy International Airport and LaGuardia.

From a temporal perspective, most human beings are active during the daytime and sleep at night. Not surprisingly, more geo-tagged tweets appear during the daytime. For example, Figure 31 shows four maps of geo-tweets throughout the continental United States, each of which represent all the geo-tagged tweets in two-hour time periods throughout a 24-hour period. Apparently, in Figure 31(a), while people on the West Coast are getting ready for bed, those in the East Coast have already fallen into a deep sleep, so the West is brighter than the East, in Figure 31(b). While people on the West Coast are still asleep, those in the East Coast are just starting their day, so, at this time, the West is dimmer than the East. In addition, comparing Figure 31(a) and (b), we can see an obvious transition from bright to dim in the West Coast, as well as another transition from dim to bright in the East Coast.

This figure illustrates how, time is another crucial feature that affects the number of geo-tagged tweets around the world.

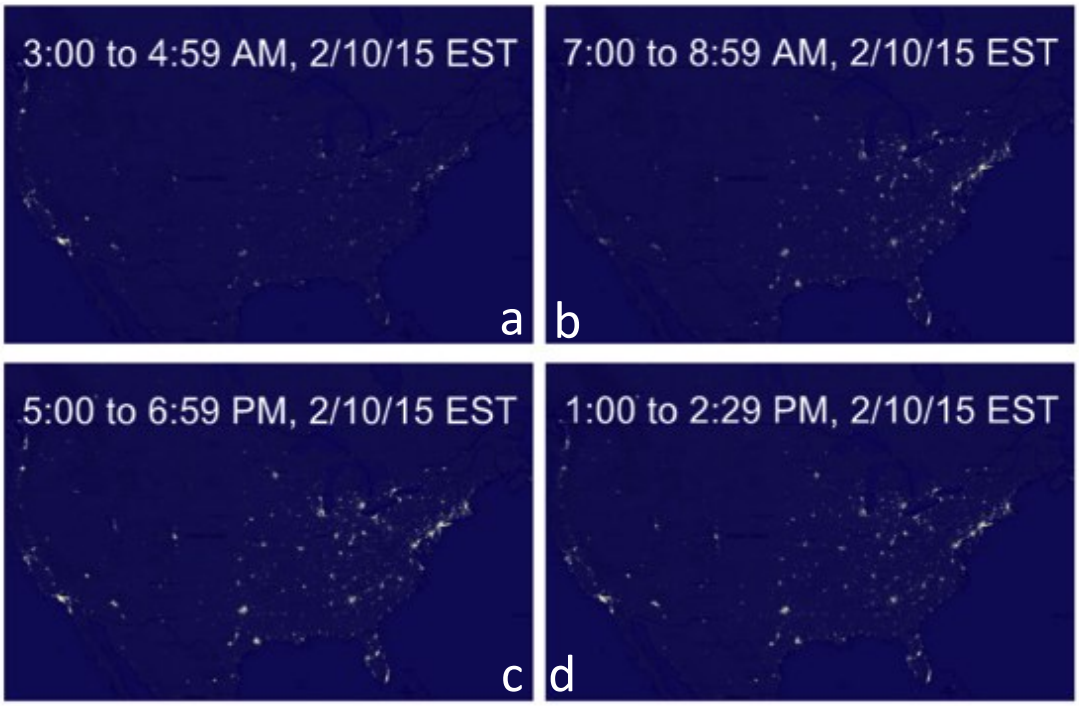

Figure 31. The temporal dimension of the geo-tagged tweets over the continental United States

From a contextual perspective, I examined the "source" property of the tweets. This property indicates the client or platform by which Twitter send the tweet. It is worth noting that, in addition to Twitter official apps, users can tweet via third-party social media apps. These sources rely on Twitter API, which is able to synchronize a user's message directly to the Twitter server. Figure 32 shows how to repost a geo-tagged photo via Instagram. When a user selects the Twitter icon, this photo can be directly synchronized with the

Twitter service. In other words, the "source" property can imply: (1) the type of device in use (e.g., iPhone, Android, iPad, Windows Phone, Blackberry, etc.), (2) the type of app: official ones developed by Twitter, or those of a third-party (e.g., Foursquare, Instagram, Path, Flickr, etc.), and (3) the general usage of the app. In this dataset, 942 sources were identified and the total number of tweets from a specific source was found to range from millions to merely a few. Almost 90 percent of all tweets are sent out through two types of sources, including official twitter clients (e.g., Twitter for iPhone, Android, Windows Phone, Blackberry, etc.) and third-party apps for personal use (Foursquare, Instagram, Path, Flickr, etc.). The author uses the geo-tagged tweet to indicate where his or her activities take place; therefore, I refer to these tweets as individual-based accounts.

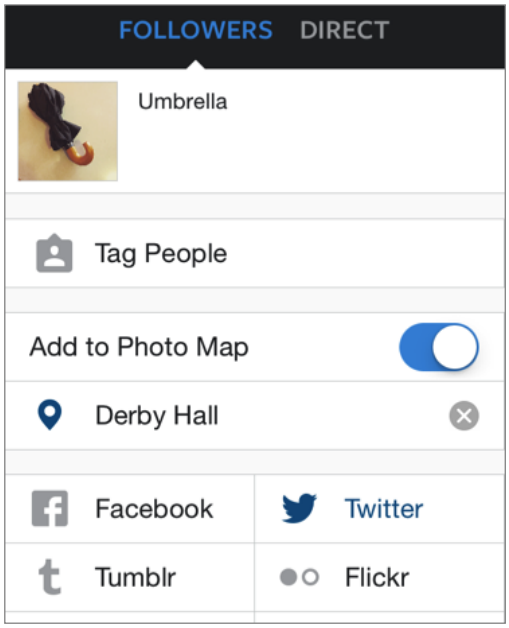

Figure 32. Tweet reposting via Instagram

Source	# Tweets	Description
Twitter for iPhone	1,056,878	Official Twitter app for iPhone
Twitter for Android	935,859	Official Twitter app for Android
Instagram	88,292	Photo and video sharing
Foursquare	70,877	Local search and discovery
Path	47,244	Location based diary
Twitter for Andrd Tablets	37,955	Official Twitter app for Android Tablets
Twitter for iPad	37,198	Official Twitter app for iPad
Twitter for WinPhone	16,170	Official Twitter app for windows Phone
TweetMyJOBS	14,755	Secure online job posting
dlvr.it	7,169	Sharing contents across multiple social media
Sandaysoft Cumulus	7,000	Weather updates at regular intervals
Twitter for Android	5,055	Another version of official Android client
Tweetbot for iOS	4,460	Tweeting on Mac
TweetCaster for Android	3,462	An alternative to tweeting
NightfoxDuo iOS	3,193	A game using Twitter for social networking
iOS	3,009	An iOS integrated function for tweeting
Untappd	2,471	Sharing locations of breweries and bars
TweetMyJOBS	2,448	A social and mobile job distribution platform
tenki	1,273	Weather reports in Japan
twitterfeed	1,181	Tweeting updates from a blog

Table 1. Top 20 sources measured by the number of tweets

Of course, Twitter is not only for personal use. There are also non-individual based accounts, including advertising companies (location based advertising about jobs, products, and services. e.g., @LoansHomes, @IvoryStandard, @fabbing), NGOs (e.g., disaster forecasts and warnings @tenkjp, @SemanticEarth, @QuakeSOS), government/military agencies (traffic accidents/traffic jam warnings, emergency notifications, recruitment, e.g., @cyber, @USArmy, @pdx911police) and robots (testing, debugging, and monitoring, e.g., @MarsBots, @GooGuns). The above-mentioned use of geo-tags is reasonable in a

geographic sense, even if these geo-tags might not follow the geographic constraints proposed by Hägerstrand (see Section 2.3). Twitter has ensured some degree of reliability to the non-individual based accounts through a trust mechanism. It allows these accounts to geo-tag from customized sources, but at the same time, severely punishes (by e.g., suspending the services or even blocking the Twitter accounts) the owner of a source if spoofing or fraud[13] is discovered. Therefore, no one would risk such a penalty by purposely designing a source for spoofing. Hence, in this dissertation, I assert that the identified trust mechanism can significantly decrease the occurrence of spoofing from non-individual accounts.

As shown in Table 1, the majority of the harvested geo-tagged tweets are all individual-based accounts. The top four tweet numbers are official twitter apps for iPhone and Android, Foursquare, and Instagram. These account for over 94.37% of all tweets. Therefore, the empirical study mainly focuses on screening geo-tagged tweets that have been sent from these four major sources.

4.1.2 Variables for detection

Variables for detection may vary according to different empirical studies. Hence, this section aims to determine two key variables for implementing this detection: (1) the speed

[13] Refer to https://support.twitter.com/groups/56-policies-violations#topic_238

of travel and (2) the threshold by which to differentiate the crowded places from the sparse.

- Speed of travel

Most long-distance flights are through commercial airlines. The maximum cruise speed for the major commercial planes, such as Boeing, Airbus, McDonnell, is approximately 1,000 km/h[14]. Therefore, I can deduce that the maximum speed of a person traveling by air Va_{max} would be no greater than this cruise speed.

In addition to travelling by air, another major means of transportation is by ground vehicles, such as cars, trains, and high rails. For any person travelling on the ground, the average speed cannot exceed the maximum ground speed Vt_{max} and reach the maximum air speed Va_{max}. In other words, with regard to travel between two geo-tags, if its average speed V is between Vt_{max} and Va_{max}, then it must be a fake path. Even though, I cannot be completely certain that any path of this speed is fake since the average speed V can be any value from 0 to Va_{max} if the user does not post tweets immediately before taking off or after landing. So, if the actual speed is between Vt_{max} and Va_{max}, there is a small chance that the path is fake. Hence, constraining conditions are required to ensure at least part of the given path is travelled by air. Specifically, since it is nearly impossible to send locational information on board due to little WIFI coverage and the GPS absorptive feature

[14] Refer to http://www.united.com/web/en-US/content/travel/inflight/aircraft/default.aspx

of the airplane body[15], it must fly at least 120 km Da_{min} in roughly 45 minutes[16] Ta_{min} to create a space-time path with an average speed between Vt_{max} and Va_{max}. Then, I can firmly assert p as p_s if its distance and time frame are both less than Da_{min} and Ta_{min}, its average speed is between Vt_{max} and Va_{max}.

The maximum ground vehicle speed Vt_{max} varies between countries or regions of the world. Specifically, Vt_{max} is equal to 240 km/h in North America (the maximum operating speed of Amtrak), 320 km/h in Europe (the maximum operating speed of TGV in France, ICE 3 in Germany), 320 km/h in Japan (the maximum operating speed of *Shinkansen*, literally translated to "new trunk line") and 380 km/h in China (the maximum operational speed of Harmony CRH 380A. Though Shanghai Maglev has a higher maximum operational speed of 430km/h, I ignored this fact since, in reality, it does not run at this speed for extended periods). Considering that most countries do not have such fast

[15] According to a report from *USA Today* Mutzabaugh, B. 2013. Routehappy unveils report card for in-flight Wi-Fi. USA Today. that 38 percent of domestic flights are equipped with in-flight Wi-Fi, I can infer that sending an information entity is possible on some flights. Even so, posting geo-tagged entities is still difficult, mainly due to the fact that GPS signals inside the main cabin are still weak and intermittent. Microwave-absorbing materials have been widely adopted in the construction of airplane bodies because of they are lightweight, very strong and possess a perfect property of electromagnetic wave absorption Chung, D. (2001) Electromagnetic interference shielding effectiveness of carbon materials. *carbon,* 39**,** 279-285.. Since GPS signals are typically electromagnetic waves, it is almost impossible in most circumstances to receive GPS signals on a plane,

[16] Star Alliance, as one of the leading global airline networks, has the largest number of daily departures, more than 21,900. According to its timetable, the distance and timeframe for the shortest flight is 120 km and roughly 45 minutes. See details of this timetable at http://www.staralliance.com/assets/doc/en/services/ tools-and-downloads/ pdf/StarAlliance.pdf.

trains, I used 320km/h as the default value of Vt_{max} for the rest of the countries or regions in the world. So Vt_{max} is a list of values based on travel paths.

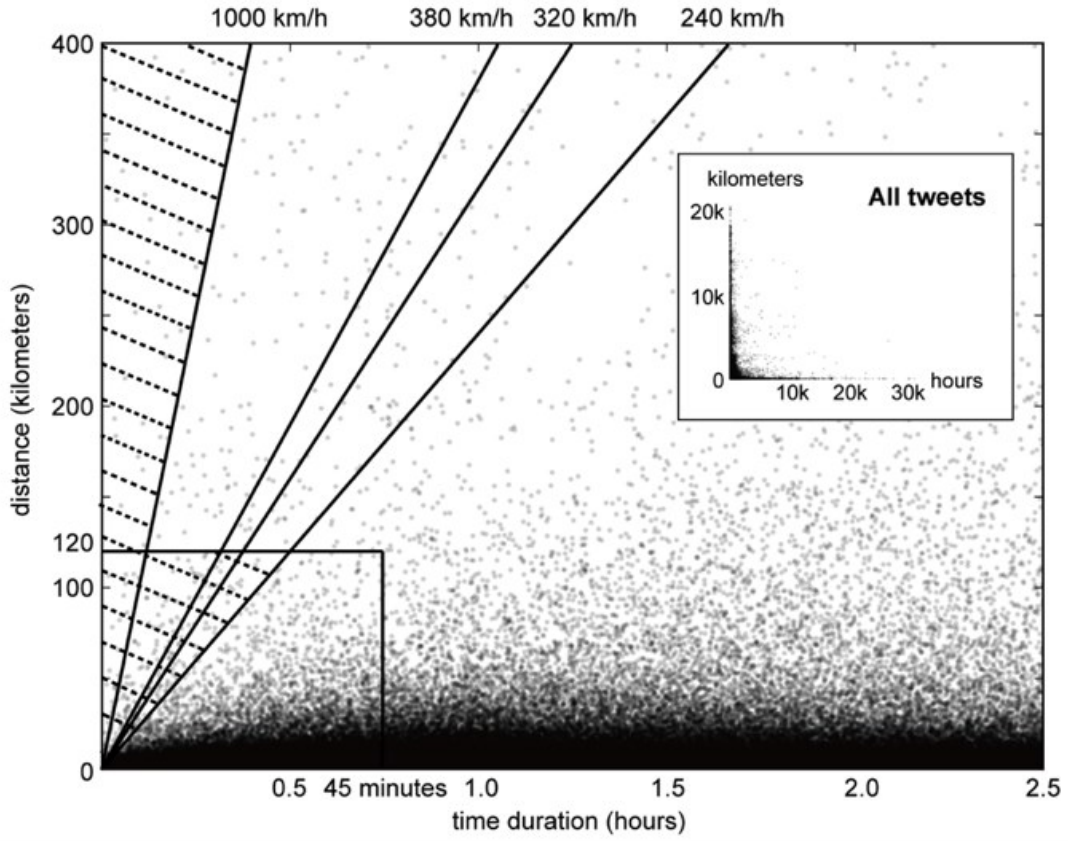

Figure 33. Geo-tagged tweets in a d-d plot

Figure 33 is a distance-duration plot (y-axis denotes distance, x-axis denotes time duration), and every dot denotes an ST path generated from the individual account. Based on the speed limits described above, the fake paths are assumed to be in the upper-left portion of the line (indicating Va_{max}, 1000 km/h) as well as the region within the boundaries made up by Va_{max} (1000 km/h), Da_{min} (120 km), Ta_{min} (45 minutes) and Vt_{max} (380 km/h

in China, 240 km/h in United States, 320 km/h in Europe, Japan and the rest of the world respectively. For example, the area denoted by the dashed line in the figure contains all the paths within the United States).

- Tweet density

As previously shown, $P(H|A)$ is a piecewise function determined by tweet density at a place l within a specific time period. Crowded and sparse places can be identified through the density D, the threshold p.

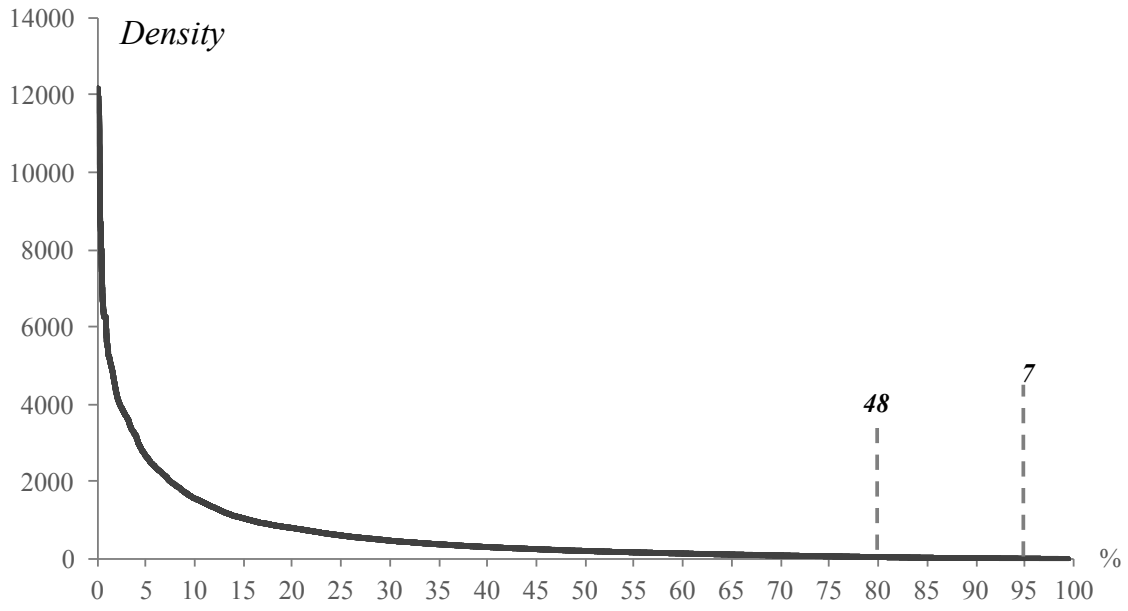

Figure 34. Distribution of tweet density

Ideally, D can be interpolated using a kernel density function. However, it would be a difficult task to calculate the millions of tweets in the empirical study. Therefore, in order

to minimize the computational complexity, I proposed an expedient way to formulate the density of geo-tagged tweets D at any arbitrary location l, that is, to count the total number of tweets within a fixed size window located at l during a specific period (e.g., an hour, a day or etc). Here, the density would be approximately equal to the total number of tweets within a particular window. This formula makes it much easier to figure out the threshold p.

Therefore, I set the window as a 5*5 km^2 square, mainly because it is an easily walkable area for the average person. Then, I randomly chose 10,000 locations and enumerated the total number of geo-tweets within the square during the two-day period (from Feb 9 to 11, 2015). Consequently, I was able to generate a distribution of the counts in each window. Figure 34 shows the distribution of the density and the results of the two different options for the threshold: The first option indicates that the top 80 percent of the 10,000 random places were crowded and the rest were sparse. If this were the case, the threshold would be 48 geo-tags within a 5*5 km^2 square during the two-day period. The second option indicates that 95 percent of the places were crowded, and, in this case the threshold was seven geo-tags. When I compared the real situation on the ground at various places with these two different numbers of tweets, I finally chose the second option.

4.2 Results

After inputting all the required variables into the algorithm, I found that the proposed detecting approach could successfully sift out fake geo-tags that are inconsistent with human constraints in space-time. Table 2 shows a summary of the detected fake geo-tags. In general, the fake ones only accounted for 0.22 percent of the entire sample. For those fake geo-tags from official Twitter clients, such as iPhone or Android, the percentages were even lower. Interestingly, though fewer geo-tagged tweets were sent from those third-party apps, the percentages of fake ones were much higher (around an order of magnitude) than those from the official Twitter apps.

Source	# Fake	# Total	% Fake in total
Twitter for Android	1,456	935,859	0.16%
Instagram	1,199	88,292	1.36%
Foursquare	1,140	70,877	1.61%
Twitter for iPhone	990	1,056,878	0.09%
Total	4,785	2,151,906	0.22%

Table 2. Detected fake geo-tags

After obtaining a general idea about how this spoofing phenomenon affects the overall data quality, I continued to analyze its spatio-temporal dynamics. To do so, I categorized the raw data divided into three groups, the original geo-tags, the detected fake geo-tags, and the cleaned geo-tags (the original data set removes the fake ones).

4.2.1 Temporal preferences

First, I analyzed the temporal preferences for sending fake geo-tags. For each tweet in the data set, the timestamp attribute was in Greenwich Mean Time (GMT). Due to the correlation between the local time and daily habits, such as the time for getting-up, working, commuting, sleeping and so on, I converted the timestamp to the local time, and then counted the total number of tweets per hour. As a result, Figure 35 shows the hourly rate at which geo-tags are posted each day. The fake geo-tag appeared more frequently during 20:00 to 23:00 o'clock, and less frequently from 2:00 to 6:00 o'clock. When compared to the posting of original and cleaned ones, more fakes appeared during the daytime, particularly, during the morning rush hour (from 7:00 to 9:00 o'clock, the period when people get up, commute, begin the workday) as well as during lunchtime (13:00 to 15:00 o'clock). This activity was less likely during the night (from 21:00 to 4:00 o'clock of the next day). Apparently, the temporal preference for sending fake geo-tags was quite different from the usual. A possible explanation is that Twitter users are more likely to send out fake locational information when travelling (e.g., during the morning rush hour and at lunchtime); therefore, they are less likely to do this when staying relatively still (e.g., during the working or sleeping hours). Even so, the fake geo-tags did not greatly affect the temporal characteristics of the overall data quality. Just as observed, sorting out the fake geo-tags from the original data set did not change this trend dramatically (because the

trends of the original and cleaned were very similar to one another).

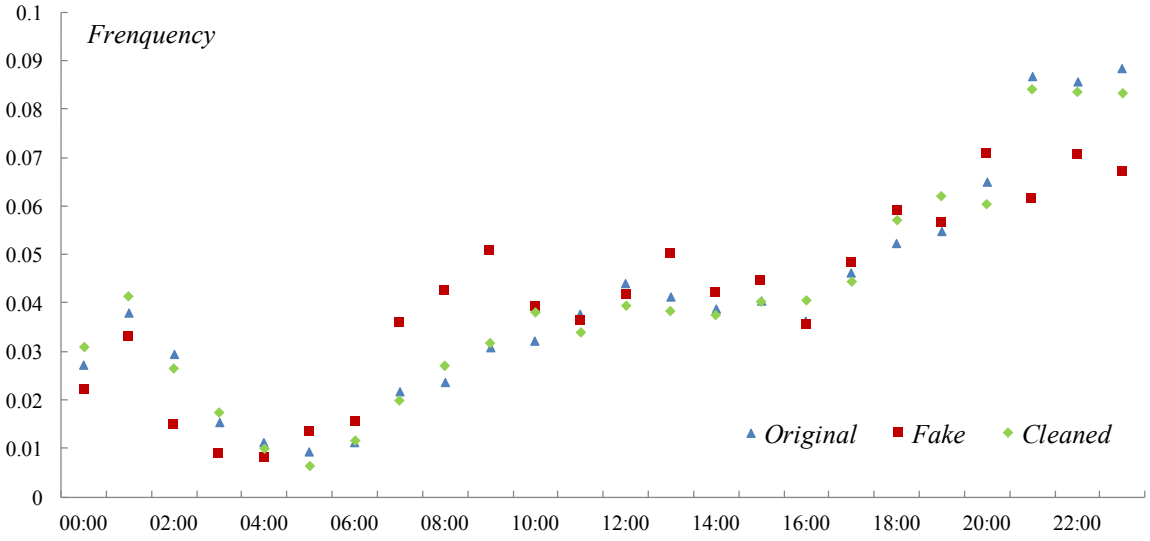

Figure 35. Frequencies of posting geo-tagged tweets

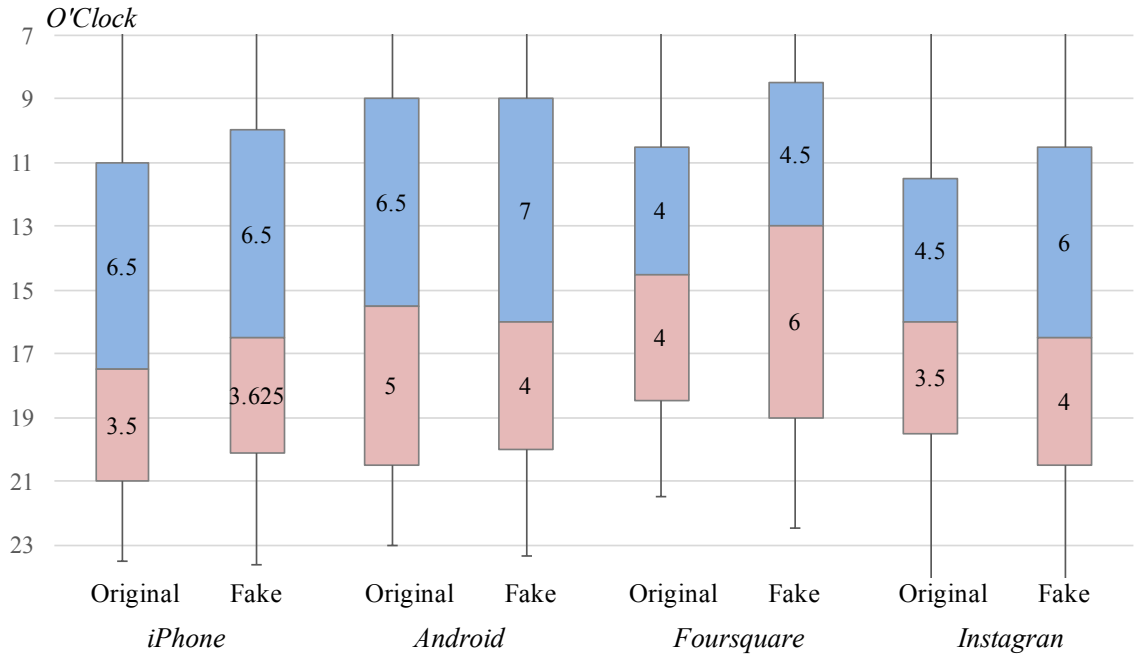

Figure 36. Temporal patterns of posting geo-tagged tweets thoughout a day

Moreover, I examined the temporal pattern of posting fake geo-tags from different sources. Figure 36 lists four pairs of box plots that represents four different sources, including iPhone, Android[17], Foursquare, and Instagram respectively. Each pair contains two boxes: while one indicates the time sequences of posting fake geo-tags, another indicates that of posting original ones. For each of the four sources, the difference between the median time of posting fake and original geo-tags were within one hour, and the differences in Foursquare were longer (approximately two hours). Moreover, the majority of geo-tags (in any box, the geo-tags, occurring from the beginning of Q2 [the blue stripe] to the end of Q3 [the red stripe], account for 75 percent of the entire number of geo-tags in that box), the differences in time interval between the fake geo-tags and the corresponding original ones were not significant for iPhone and Android, whereas the differences in time interval were longer for Foursquare and Instagram. For example, most fake geo-tagged tweets from iPhone were sent from 10:00 to 20:00 o'clock, while the original ones from iPhone were from 11:000 to 21:00 o'clock, both the time intervals were 10 hours. Moreover, most fake tweets sent from Foursquare were between 8:30 to 21:00 o'clock, while most original ones were sent from 10:30 to 18:30 o'clock. The difference in time interval was approximately two and a half hours. Overall, in terms of temporal characteristics, the tweeting and spoofing activities were found to vary between sources/platforms. Also, the time

[17] In the following paragraph, iPhone and Android indicate the official Twitter apps for iPhone and Android respectively unless notified.

sequences for posting fake geo-tags were quite different depending on the sources. This result makes sense because different sources provide specific functions and appeal to different target user groups.

4.2.2 Spatial distribution

This section investigates the spatial characteristics of the detected fake geo-tags. In Figure 37, all fake geo-tags are categorized by source and spread out globally. Consistent with the global distribution of the entire sample (refer to Figure 29), the fake geo-tags were also unevenly distributed at the global level. As observed, whilst the authors of fake geo-tags in North America and West Europe prefer to use multiple clients, Japanese and Southeast Asian authors prefer Android while Brazilian authors prefer Foursquare. Apparently, these various sources/platforms create patterns which reflect the usage habits and reveal that the market position varies in different regions over the world.

Location spoofing also indicates unique spatial characteristics in urban areas. Take New York City for example (see Figure 38), most fake locational information is clustered in downtown Manhattan, while less is scattered around the suburban areas. The fake geo-tags from Instagram were more clustered than the other three sources. In the suburban areas, more fake geo-tags from iPhone or Foursquare were found, and fewer were from Instagram. Furthermore, at least in New York City, Android was seldom used for sending fake geo

tags.

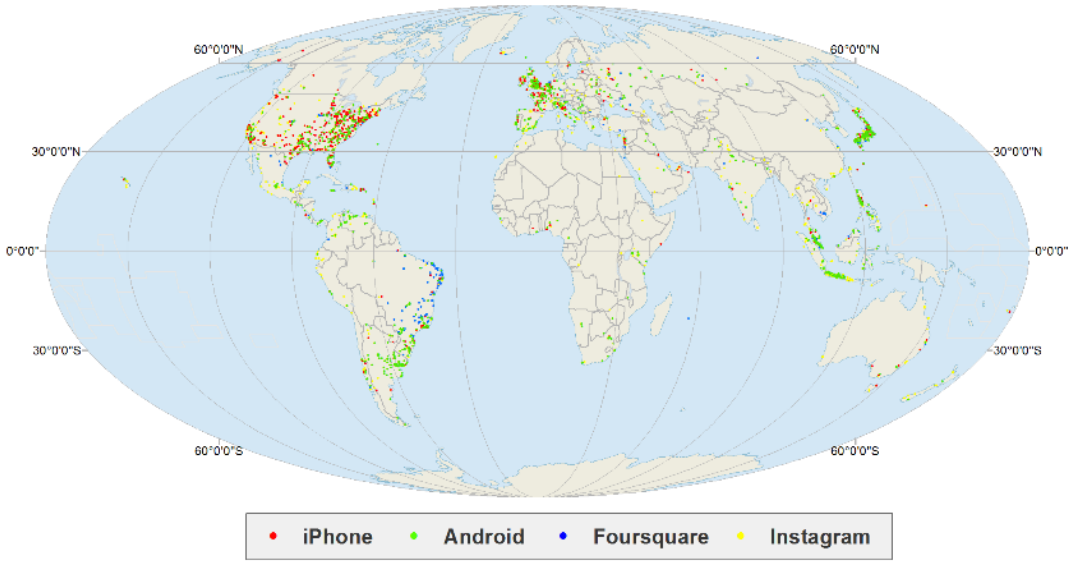

Figure 37. Global distributions of fake geo-tagged tweets

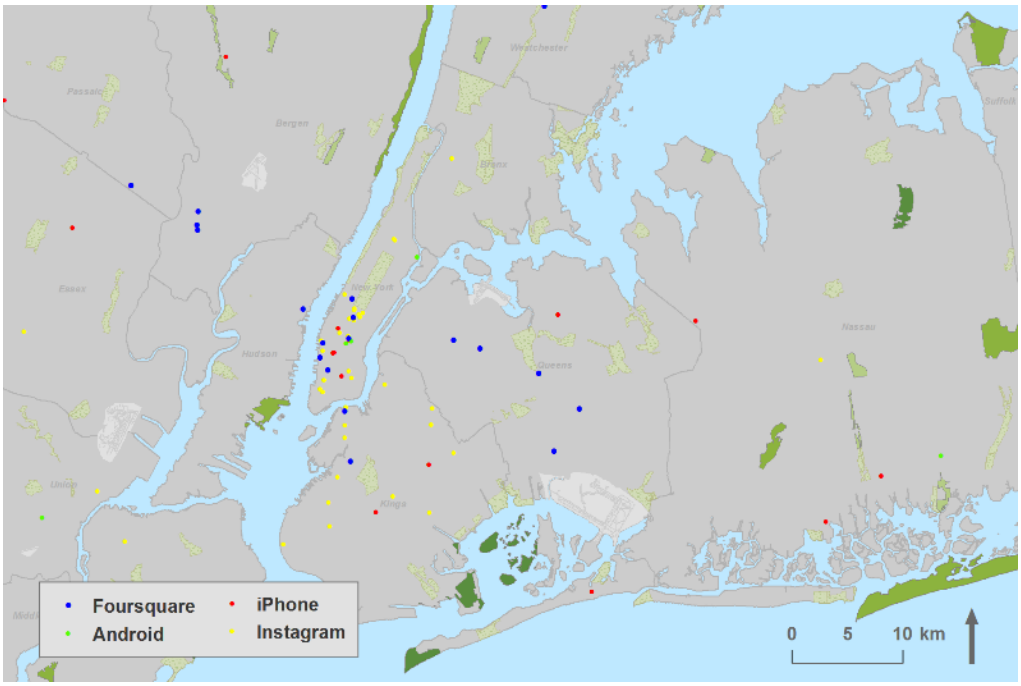

Figure 38. Fake path in New York city

	Observed Mean Dist.	Expected Mean Dist.	Nearest Neighbor Index	Z-score	P-value
Original	107,848	376,413	0.287	-42.63	0.000000
Cleaned	90,495	…	0.240	-43.59	…
Fake	99,499	…	0.264	-42.22	…
Fake iPhone	84,105	…	0.223	-44.57	…
Fake Android	86,534	…	0.229	-44.20	…
Fake Instagram	90,139	…	0.239	-43.65	…
Fake Foursquare	63,738	…	0.169	-47.67	…

Table 3. Nearest neighbor index (NNI) analysis results

Having laid out the spatial distribution of location spoofing as the foundation, I also quantitatively examined the spatial range of fake geo-tags based on the nearest neighbor analysis. In general, this analysis calls for measuring the distance between each geo-tag in the data set and that of its nearest neighbor and then averaging all these distances. The Nearest Neighbor Index (NNI) is a ratio of the observed mean distance to the expected mean distance. The expected distance is the average distance between neighbors in a hypothetical Complete Spatial Randomness (CSR) distribution. If the index is less than 1, the pattern favors clustering; if greater than 1, the pattern favors dispersion. Additionally, the statistically significant measures of z-score and p-value, suggest whether or not the distribution rejects the null hypothesis (indicating if the features are randomly distributed), and whether the features indicate statistically significant clustering or dispersion. Specifically, the p-value is a probability indicating the level of randomness of an observed spatial pattern. When the p-value is small, it is very unlikely (small probability) that the

observed spatial pattern is the result of random processes. This information rejects the null hypothesis. The z-score is the number of standard deviations when an observation is above (positive) or below (negative) the mean. Great absolute values of z-scores, associated with small p-values, are in the outskirts of normal distribution. In this situation, it is very likely that the observed spatial pattern rejects the theoretical random pattern, and instead shows a cluster pattern.

Table 3 lists the results of applying the nearest neighbor analysis to various groups of geo-tags. To make the results comparable, 900 geo-tags were randomly distributed throughout each group. Given their z-scores from -42 to -48, all the listed geo-tag groups were not randomly distributed. Moreover, among the original, fake, and cleaned samples, the fake geo-tag sample revealed the largest NNI value, whereas the cleaned sample had the smallest. This result may imply that, overall, the fake geo-tags have a wider distribution pattern are less clustered compared to the normal geo-tags (the cleaned sample). Furthermore, the clustering level varied between sources, and the fake geo-tags sent via foursquare platform were much clustered than those sent from other sources/platforms.

NNI is a global variable that indicates the level of spatial clustering, but it falls short in describing the local distinctions. Thus, I investigated the frequency of nearest neighbor distances for various lengths. Figure 39 shows the cumulative frequencies of nearest

neighbor distances for the original, cleaned and fake geo-tags. Likewise, to avoid the impact of density, each group equally sampled 900 randomly distributed geo-tags. With regard to a specific group of geo-tags, the cumulative frequency indicates the likelihood that the nearest neighbor distance equals a certain length, it further reveals whether the specific group of geo-tags is more likely to appear in an urban cluster or in remote regions. It must be noted that approximately 90 percent of each of the three frequency trend samples were clustered within 200 kilometers. The tendencies of both the original and cleaned samples were approximately the same; however, the fake samples differed significantly. When the distance was less than 20 kilometers, the trend of fake geo-tags was a little steeper, when over 20 kilometers, the trend of the original sample was slightly higher than that of the cleaned tweets. This comparison indicated that eliminating the detected fake geo-tags did not severely affect the spatial cluster pattern. In contrast, the trend of fake geo-tags indicated a unique tendency. As this trend was steep when the distance was less than 15 kilometers and became relatively flatter, when the distance was longer than 160 kilometers. This curve became steep again and surpassed the other two trends when the distance reached 180 kilometers. The trend of fake geo-tags indicated that, the fake geo-tags were more likely to occur in urbanized areas (no more than 15 kilometers) as well as remote areas (of more than 180 kilometers).

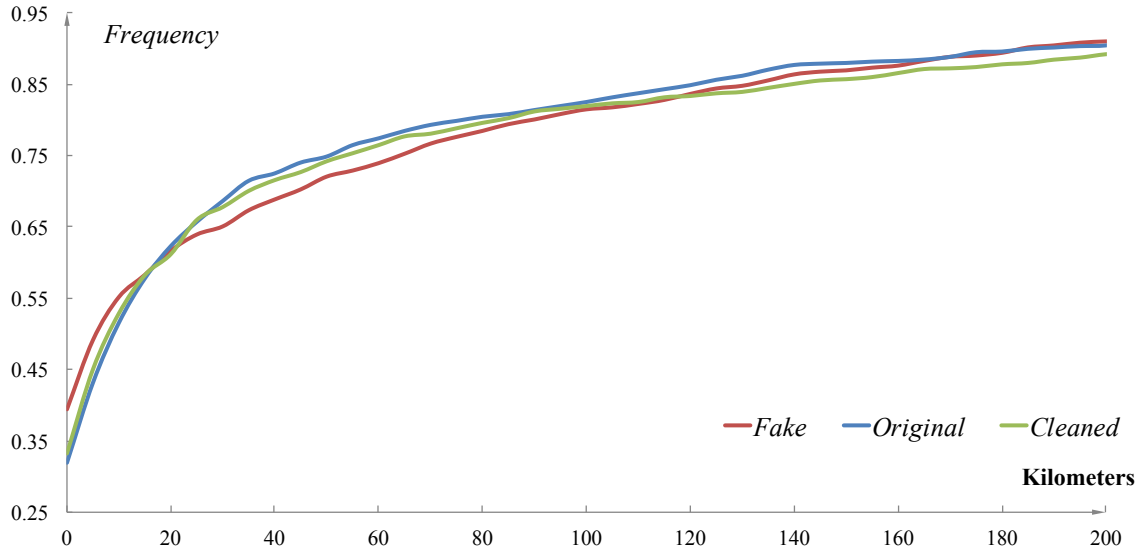

Figure 39. Cumulative frequencies of nearest neighbor distances

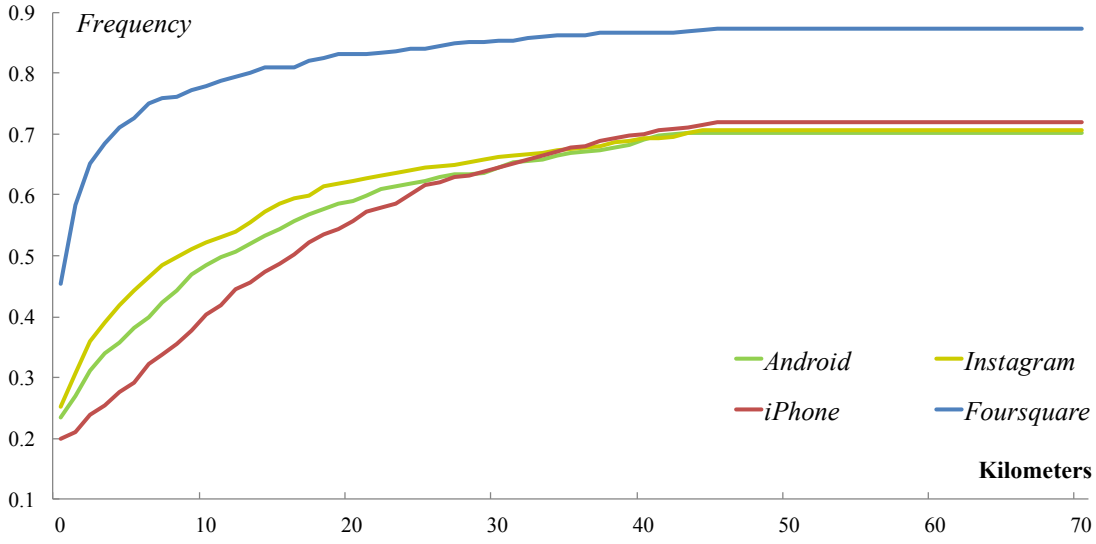

Figure 40. Cumulative frequencies of nearest neighbor distances of differnt sources

Furthermore, I investigated the spatial patterns of fake geo-tags sent via the four major

sources. As shown in Figure 40, more than 60% of the fake geo-tags from iPhone, Android and Instagram as well as more than 85% of the fake geo-tags from Foursquare were within the range of 40 kilometers. According to the slope of the four curves, the majority of fake geo-tags, no matter which source was used, were more likely to appear in an urban area (a region with a radius approximately 20 kilometers, equivalent to the size of the core area of an average county). Moreover, among these four platforms, the fake geo-tags via Foursquare were even more clustered in the urban area.

4.3 Evaluations

To better grasp the utility of the detecting approach and understand its limitations, I evaluated it based on two major criteria: the time variation and the distance between consecutive tweets sent by the same suspected location spoofer.

Figure 41 illustrates the time interval between a space-time path (a path linked by a geo-tag and the consecutive geo-tag sent by the same Twitter user). The three curves respectively illustrate the frequencies of the time intervals of original paths, fake paths, and cleaned paths (the original paths rule out the fake ones). To smooth the curves, I set a five-second bind for assembling data. Moreover, considering that approximately 50 percent of the original and cleaned geo-tags as well as 90 percent of the fake geo-tags were sent within

the first 16 minutes, only the main portion of these frequency curves (the time variation of the curves from origin to 16 minutes) are displayed. As shown, these three curves reached their peak at approximately the 15-second mark. But the difference was, the peak of the fake-path curve was much higher than the other two, considering that the tendencies of the original and cleaned were quite similar to one another. This fact actually implied that ruling out the fake geo-tags from the original ones dids not remarkably improve the overall data quality. After they peak, the three curves all sharply decline after two to three minutes and then became even and flat. This comparison indicated the time interval between posting two consecutive geo-tagged tweets varied from short time intervals, such as seconds, minutes, and hours, to long time periods, such as days, months, and even years. (To better demonstrate the intricacies of the distributions, this figure does not show the curves of time intervals longer than 16 hours. However, the long time interval was explored in this data set). Moreover, according to the original geo-tag trends, Twitter users were more likely to post continuous geo-tagged tweets in a short time period of 15 seconds, and most geo-tagged tweets were created within two or three minutes.

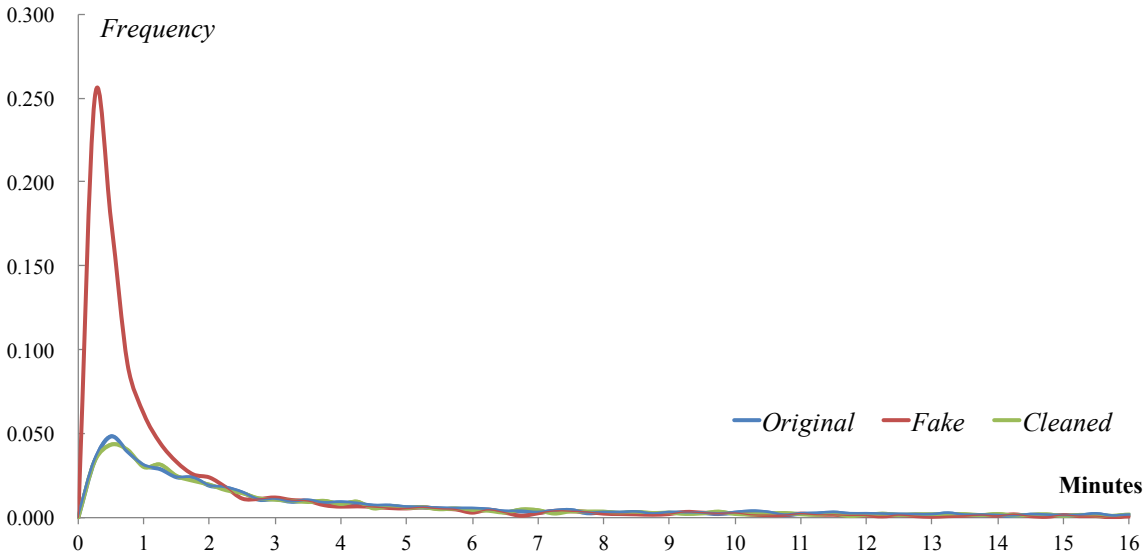

Figure 41. Frequencies of the time interval between posting consecutive geo-tagged tweets

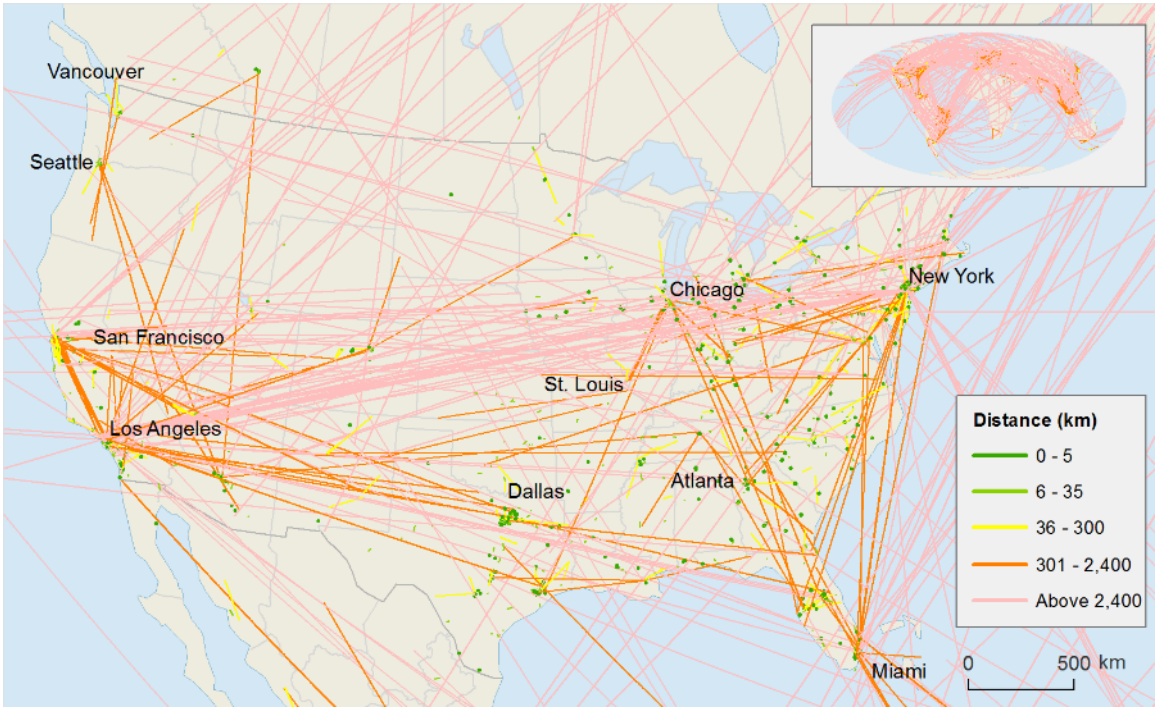

Figure 42. Fake paths of location spoofing over the continental United States

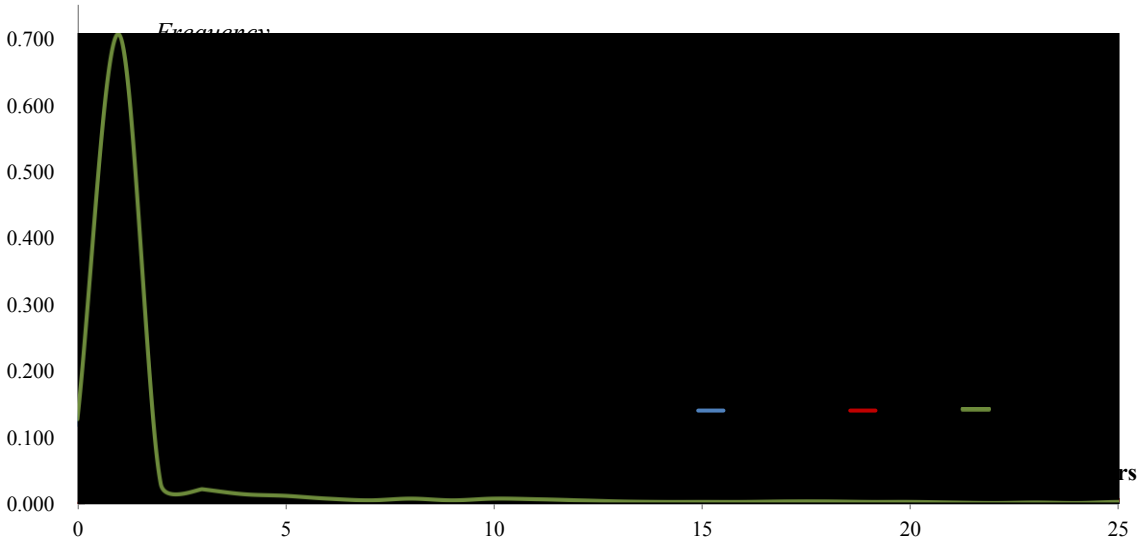

Figure 43. Frequencies of distance between two consecutive geo-tagged tweets

Additionally, I analyzed the distance at which Twitter users were likely to spoof a geo-tag. Figure 42 shows all fake space-time paths over the continental United States, and the inset map shows fake paths around the world. Each line denotes a fake space-time path linking two consecutive geo-tags, at least one of which is spoofed. I separated the fake paths into five groups using a geometric quantile classification. Not surprisingly, I found that, some metropolitan areas like New York, Los Angeles, San Francisco, Chicago and Miami were the major sources and destinations of location spoofing. Also, the fake paths were different lengths and crossed cities, states, and even continents. Considering the interstate fake paths for example, there were many fake paths between New York and Los Angeles, Miami, Tampa or Atlanta, between Vancouver and Seattle, between Chicago and St. Louis or Los Angeles, and between San Francisco and Los Angeles. Overall, this finding demonstrates

that people may falsify their locational information to either nearby or far-away places. Viewed on a national level, the long-distance location spoofing mainly occurred in major metropolitan areas.

In addition to visualizing the "spoofing distance", I also studied it in a more quantitative manner. Figure 43 shows the frequency of "spoofing distances" of three different geo-tag groups. Notably, to smooth the curves without losing the general trends, the limit for assembling data was specified at one kilometer. Also, considering that approximately 83 percent of original and cleaned geo-tags as well as 62 percent of fake geo-tags are sent within 25 kilometers, I only determined the frequencies in the 0 to 25 kilometer range. As shown, both the original path and cleaned path had very similar tendencies. Since it is possible to send out continuous tweets from one location, these two curves began to expand from some point on the y-axis, rather than the origin point, and rapidly increased to a peak at approximately one kilometer. After reaching a peak, the curves dropped down dramatically at two kilometers, and then flattened. As for the curves of the detected fake paths, they began from the origin and climbed to a peak at approximately 2.5 kilometers. Then, then slowly descended and then flatted. Notably, from two to approximately 15 kilometers, this curve of fake paths became much higher than the other two. This result implies, on one hand, the inefficiency of the proposed approach in detecting fake geo-tags at the same locations or a small region within a radius of approximately one kilometer. On

the other hand, this method is much effective for discovering fake geo-tags that are being spoofed to a region with a radius of one to 15 kilometers, which is approximately the size of an intra-city area.

Overall, through comparison of the original, fake and cleaned sample, I tested the utility of the proposed approach. I found that this approach is capable of detecting certain types of location spoofing from large-scale social media feeds, sifting out the fake locational information created within a short time period of approximately two to three minutes or falsified to an intra-city area with a radius ranging from one to 15 kilometers. However, admittedly, this approach can by no means detect all types of location spoofing partly because it is only affected by the three human constraints in space-time (especially the physical mobile constraints for individuals). To increase the detecting rate, this approach may require more geographic knowledge of locational inconsistency to be modeled as well as other facts about spoofing. In this sense, this approach was initially designed as a fundamental methodological hub to integrate various "plugins" about locational inconsistency.

4.4 Discussion

Having detected various location spoofing cases using geo-tagged tweets, I further

examined the hidden motivations for why location spoofers falsify their locational information and any other possible generative mechanisms. Finally, as an emerging topic in social media, location spoofing has been largely ignored by society, and the consequences of not paying enough attention to it can be profound and even deadly in some cases. To develop a holistic understanding about location spoofing in social media, I examined location spoofing under the broader context of McLuhan's laws of media.

4.4.1 Motivations for spoofing

This section further examines the motivations of location spoofers. Since locational information can be falsified during the entire course of information transmission rather than just at the critical step of pressing the "send" button, I identified three primary types of location spoofers, including, but not limited to, individual social media users, social media service providers, and government/military agencies. Their spoofing motivations, behaviors, and the concomitant accountabilities warrant further discussion.

- Individual social media users

It is commonly believed that conducting location spoofing seems impossible for ordinary Internet users, partly because prevailing social media apps do not provide any particular functions or publicly available "backdoor" programs (e.g., Back Orifice, a controversial program designed for remote system administration) to generate fake locational

information. Moreover, to create, update and modify geo-tags often requires both the external positioning systems (e.g., GPS, Wireless Local Area Network, Cell towers when outdoors, Wi-Fi, Bluetooth, while RFID and iBeacon when indoors) and the internal positioning receivers/chips embedded in desktop or mobile devices. Usually, only those social media service providers can store, manipulate and analyze the entire data set, and government agencies and military personnel are believed to use backdoor programs (as backed by the Patriot Act in the U.S.) to access social media databases under the guise of defending against attacks from cyber terrorists or protecting national security (Savage, Wyatt and Baker 2013). In spite of these challenges, as previously stated, a large group of the detected location spoofers are individual users.

The institutional and technical obstacles identified here would seem to prevent ordinary individuals from falsifying locational information. However, as a matter of fact, under the Free/Open GIS movement during the past ten years or so (Sui 2014), techniques for location spoofing, once exclusively dominated by government/military agencies and big corporations, have become more accessible to ordinary individuals. Apps for modifying locational information are publicly available for desktop browsers (Drager 2011) and mobile devices (e.g., iPhone, Android phone, Windows phone, BlackBerry, etc.). The Google Play Store, found on Android, hosts many inexpensive (below ten dollars) and even free “Fake Location” apps for location spoofing (see Figure 44). The primary

function of these apps, in general, is to temporarily block the embedded positioning receiver/chip and then send fake locational information back to the apps as geo-tags, check-ins or tracking routes. Thus, ordinary users, even without any professional skills, can become spoofers.

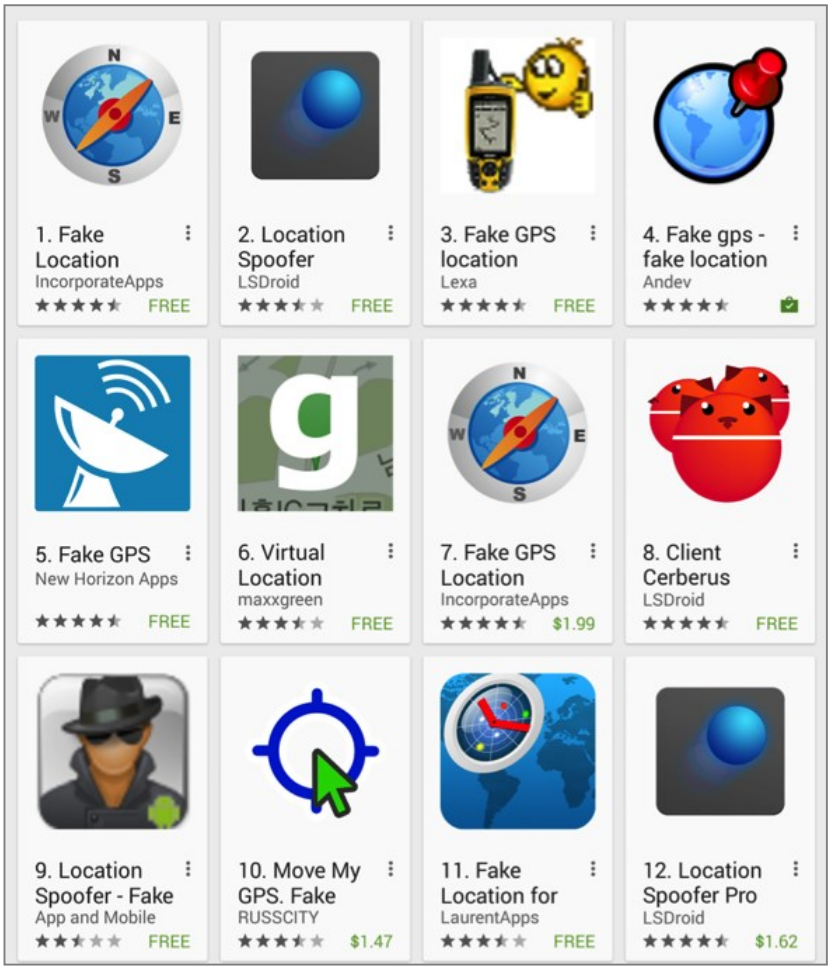

Figure 44. Location spoofing apps for Android phones

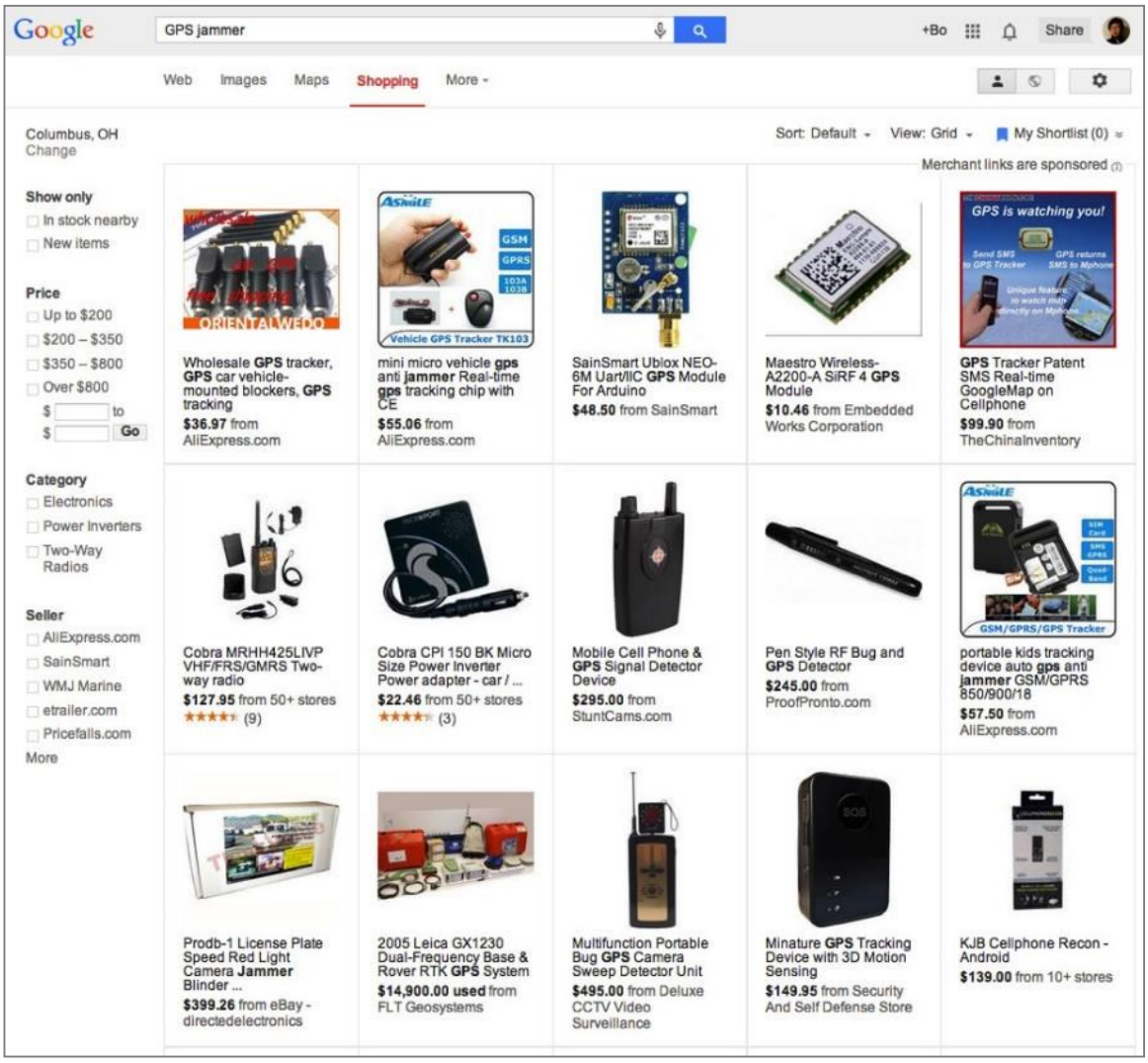

Figure 45. GPS jammers/spoofers in Google Shop[18]

In addition to software-based spoofing methods, the recent emergence of the open source hardware movement requires inexpensive electronic parts as well as publicly available hands-on tutorials and blueprints; therefore, hackers can effortlessly put together a GPS jammer/spoofer (http://www.ladyada.net/make/wavebubble). These devices can disturb

[18] Refer to https://www.google.com/search?q=GPS+jammer&tbm=shop, the web page was shot on May 31, 2014

GPS signals at various ranges, from as small as several square meters to as large as a medium-sized city (Brewin 2013). Due to their potential to cause serious disruptions, the U.S. Federal Communications Commission outlawed the use of GPS jammers in February 2012. It is currently illegal to use, market, manufacture, or sell GPS jammers (FCC 2012) – However, during the course of this study, I found that it was still possible to buy a GPS jammer/spoofer from Google Shop, eBay or Alibaba before mid-2014 (see Figure 45). Thus, an average person can easily falsify his or her locational information with these free/quasi-free apps and open devices.

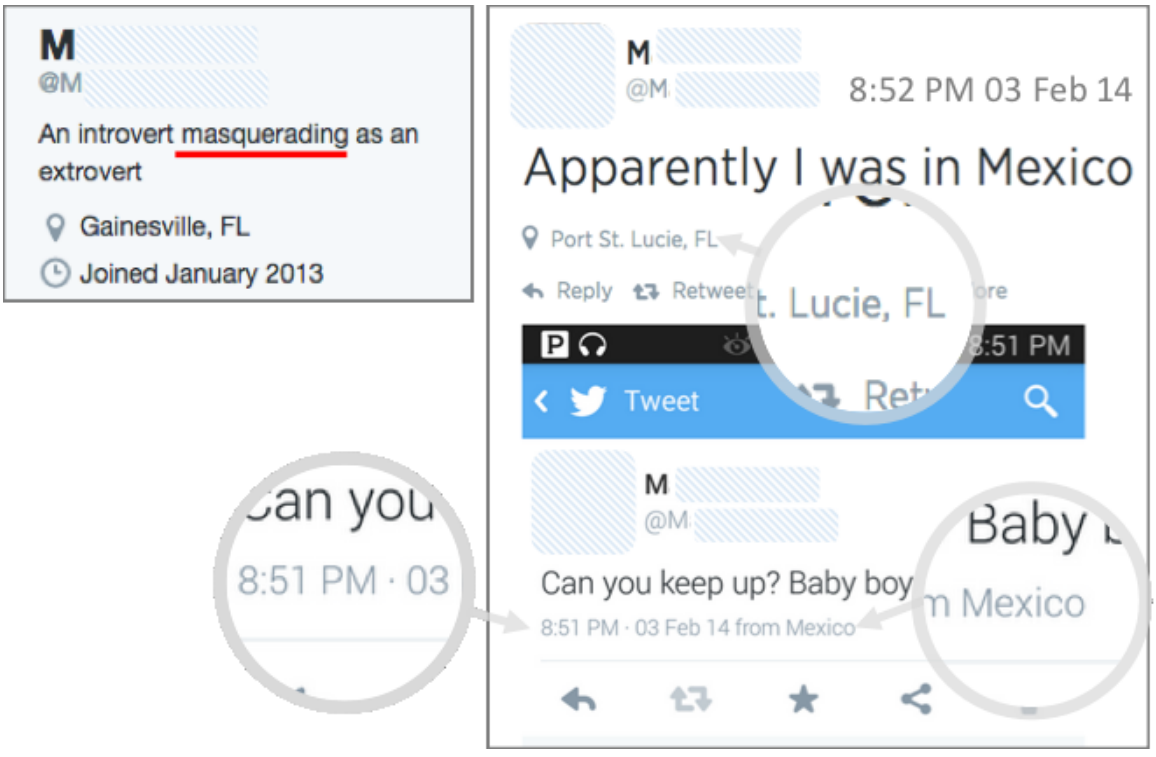

Figure 46. Location spoofing for fun

With these spoofing tools, individual social media users are capable of falsifying locational information on line. The rest of this section details a few spoofing cases from the empirical study.

Figure 46 shows a tweet attached to a screenshot of another previous tweet. This tweet was sent at 8:52PM from Port St Lucie, Florida, U.S., while the attached tweet was sent at 8:51PM from Mexico. Although this Twitter user claimed that "Apparently I was in Mexico", it is impossible for this user, regardless of the selected transportation means, to travel such a long distance from Florida to Mexico in no more than two minutes. Moreover, since most of his tweets were sent from near Port St Lucie, Florida, it is highly likely that the geo-tag in Mexico was falsified. In his profile, he describes himself as "an introvert masquerading an extrovert". Presumably, this user is very fond of masquerading/spoofing as well. In this sense, I can assume that this instance of spoofing was done simply for fun, or just to experiment with this unusual technology.

Figure 47 shows another instance of detected location spoofing performed by a Hollywood film manager. He sent out two consecutive geo-tweets, respectively, from Fujian, China and New York City within a mere eight hours. Obviously it is impossible to travel that fast from the Eastern coast of China to the East coast of the United States. During a quick interview he explained that the geo-tweet from Fujian, China was a fake. This tweet was sent out during his layover in Vancouver, Canada. He wanted to record the places he had

been to in China as a means to kill time while waiting in the airport.

Ja Feb 10
It be so cold in NYC #layover #pitstop #ohtheplacesyoullgo @ Queens, New York
Ja Feb 10
Bar that sells panties by day? So many questions... @ 万达广场 Wanda Plaza
Interview text
Ja Apr 15
I was in NYC at that time. Pictures from China were geotagged and later uploaded on layover in Vancouver. Hope that helps!
View conversation

Figure 47. Location spoofing during a layover

2 个月前 · Majestic Elegance Punta Cana, Punta Cana, Dominican Republic
SHOUT OUT TO OUR MANAGER OF DC LOCATION
TAKING SYN INTERNATIONALLY!WE
GONNA MISS YOU BUDDY!SEE YOU SOON!!
#dominicanrepublic#puntAcana#syn #grencosciece
#GlowIndustries #IslandDyes_chp #islanddyesrva
werepresent #werockhst #werockthe DR#
ockrichmond #we rockobx# werock CHP
isaandkate, alta_tec

Figure 48. Location spoofing by a dug dealer

Figure 48 displays a geo-tagged tweet that located in Dominican Republic. This author of this tweet is a drug dealer in D.C. He said, “Shout out to our manger of DC location, taking syn internationally! [...]” This drug dealer, wearing a hoodie with the word “syn” (synthetic marijuana) printed on it, appeared to be photographed while sitting in an airplane cabin. Here, the question arises, if the location in the Dominican Republic is accurate, how could he have smuggled drugs on to the airplane? It must have been the result of some flaw in airport security. However, if the location is false, a further question remains: Why did he do that? A probable explanation is that he used location spoofing as a means to internationally advertise his product. By showing potential customers that his business was famous internationally, more customers would be attracted. Or possibly, since users can search nearby geo-tags, he might have considered it a fake geo-tag as an efficient method for attracting more customers from the Dominican Republic or other places (in which selling and buying marijuana is unlawful).

I also noticed that a female movie fan as well as travel blogger posted many geo-tagged tweets to film locations in several European countries. Including texts and photos, she was able to describe the wonders of being on location. As a result, many online visitors were attracted to her twitter page. These days, it is a common marketing strategy on social media to advertise a film by showing an average viewer’s opinion. Moreover, as soon as I added this user to my list of Twitter friends, the official Twitter account of that film began

following my account. This incident further convinced me that some possible links exist between this travel blogger and the film company. She admitted that she geo-tagged some film locations when she was actually far away. As she explained, her intention in posting fake geo-tags was to identify the place where the photo was taken afterwards. Certainty, it is absolutely reasonable to geo-tag photos as a record on social media. However, as a Twitter user who makes a living by writing a travel blog, if she occasionally falsifies her geo-tags to places she has never been for some quick money from the film marketing team, could she lose the faith of her twitter audience who still believe her travel stories? Obviously, this user's "real" experiences at those places can be called into question.

Through this spoofing activity, individual users can hide their actual whereabouts as well as confuse their friends, thereby protecting their geo-privacy. Or they simply do it for fun and to experiment with this unusual spoofing phenomenon. It could also be a way to forge locational identity. Undoubtedly, the capability to access a special location can somehow reflect a privilege, vocation or other aspects of identity, which are tantalizing. For example, only military personnel are admitted into the military zones, only rich families are able to live in upscale communities, and only the big drug lords can conduct intentional drug trafficking. In this way, the audience might be tricked into believing that the location spoofer is similar to the individuals who actually live, act or work at that location.

- Social media service providers

Another significant percentage of location spoofing comes from some third-party social media apps. Intuitively, social media providers should prevent any conditions that may lead to spoofing. Indeed, most of them claim to protect individual geo-privacy, but this measure is not only for the specific purposes, as alleged. Based on my observations, social media providers prefer to tolerate location spoofing to some degree, even more, some might benefit from this spoofing phenomenon.

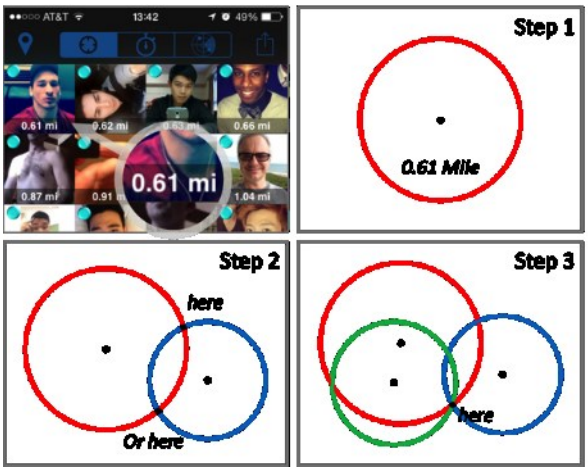

Figure 49. Trilateration used on social media

Figure 50. Location shift

A commonly used location-based function displays how far away a friend is from one's current location. For example, Jack'd, a social media app for gay chatting, displays the distances between the nearby users (see the top-right frame of Figure 49), but does not display their exact locations. However, using trilateration (a geometric measurement to

determine an unknown location by measuring the distances at three different locations, illustrated in Figure 49), a user's location can be approximately estimated. Therefore, to protect the geo-privacy of an online user, location spoofing is often adopted as a counter-measurement to degrade the quality of locational information, such as activating the location shift/spoofing function (see Figure 50) provided with a social media app "single around me"[19]. As described, this function can move the locational information approximately two to three kilometers away from one's true location without sacrificing the accuracy of location-based searches. However, the original locational information is still stored in the server to support location-based services that require an accurate location. It is a potential threat for the user to store the accurate locational information in a remote server, no matter how secure the social media service provider claims to be. A recent example is that Apple iCloud, though alleged to be a very secure cloud storage, leaked the geo-tagged nude photos of several celebrities (Hill 2014). In this sense, even if the locational data can be securely maintained, it is still possible to determine a social media user's whereabouts.

Moreover, location spoofing is used as a shrewd advertising strategy to subtly promote businesses (Boyd and Crawford 2012, Wilson 2012). Take the most commonly used function "check-in" as an example, which enables one to assign the coordinates of an

[19] This is a location based social media popular among bachelors for dating or pursuing a relationship.

actual place (e.g., restaurant, cinema, supermarket) to an online information source. Once a user is "checked-in" at any nearby location (determined by the apps), other than his or her true location, the locational information is actually falsified. For example, Figure 51 provides a partial list of the possible "check-in" places on the screen of a mobile Foursquare client, which are near the user's actual location (indicated by the blue dot in the map). As shown, the distance between one's true location and candidates ranges from 300 feet to 0.2 miles. By doing this, while spurring the local business to offer offline services (e.g., coupons, advertisings, sales) to social media users who have visited virtually, this location-based function also encourages potential customers to 'check in' at more local stores that provide offline promotions.

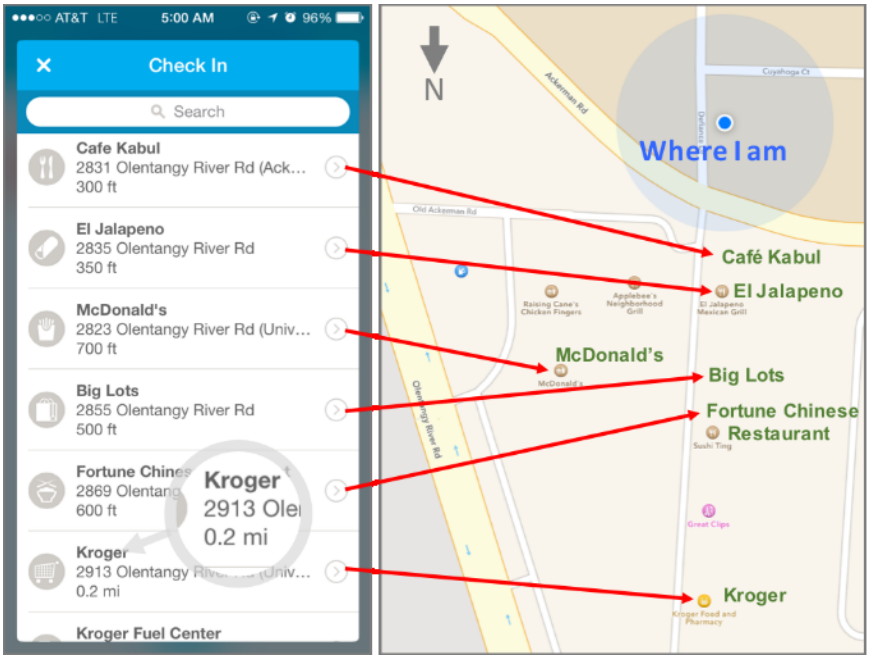

Figure 51. Location spoofing as an advertising strategy

Furthermore, Instagram, as one major location spoofing platform, has adopted a much more open attitude regarding the use of locational information. Unlike Twitter, Instagram focuses on sharing photos and short videos rather than short texts. Until May of 2015, there were more active users on Instagram than Twitter (Keach 2014). Instagram allows users to pin a photo or short video to an arbitrary location even as far as 20 miles away, which is much further than the distance range of Foursquare. Even more, it enables users to name any place that strikes their fancy. For example, Goodchild and Li (2012) have questioned the accuracy of a geo-tagged photo in a park, mainly because the tagged photo included a café named as "Joe's Coffee" which is nowhere near that park. Although, this geo-tag did not make any sense for general audience, it probably brought joy to the user who created it. Accordingly, this user can geo-tag that place as "Joe's Coffee", or even use any other name that has personal meaning to the user. Indeed, this feature gives a great deal of latitude for social media users to operate locational information. However, without a strict naming policy, VGI generated on Instagram is negatively contaminated by this arbitrary naming function. Anyone who seriously wishes to use locational data from Instagram would be wise to take into account the existence of location spoofing.

- Government and military agencies

The government and military agencies have also been acutely aware of the social-economic and political power of locational information. However, seldom have

laws or polices been issued to directly deal with location spoofing in social media – this seeming indifference does not mean the spoofing phenomenon has been ignored by government or military agencies. As a matter of fact, they are deeply involved in creating, detecting and fighting against various forms of location spoofing in social media. On October 4, 2014, Zabiullah Mujahid, a Taliban spokesman, posted a series of geo-tagged tweets (at Sindh, Pakistan) via his Twitter account @zabihmujahid (see Figure 52). Several major Western media outlets joked that he had accidentally revealed the hidden base (Press 2014). Later Mujahid sent another tweet describing this location as an "enemy plot" and leaving his Afghan telephone number as proof of not taking refuge in Pakistan. Since no further evidence was found, Mujahid might have used fake geo-tags to deceive the Western media.

Figure 52. Controversial geo-tags dent by Taliban dpokesman

Likewise, the BBC (2014) reported that the U.S. Agency for International Development (USAID, a U.S. government aid organization) is behind a social media service Zunzuneo (Cuba's equivalent of Twitter). This app "is allegedly designed to foment unrest in Cuba". As reported, Zunzuneo deliberately falsified locational information to organize "smart mobs" - mass gatherings to trigger a Cuban Spring" (Ford 2014). Undeniably, with full access and control of the database, this U.S. agency is capable of falsifying locational information if it so chooses.

4.4.2 Controversial aspects of location spoofing

According to the definition, location spoofing must be driven by certain human motivations. If it is difficult to trace, even without any direct human intervention, a piece of fake locational information cannot be classified as a case of location spoofing. Nevertheless, the fake locational information of this sort still confuses us frequently. In the Empirical Results Section, I successfully detected several controversial cases. Then, I aimed to discover how their generative mechanisms may influence these spoofing cases, including, but not limited to, the uncertainty of locational information, software design dilemma, and computer malfunctions.

- Locational uncertainty

The uncertainty of locational information, though also deceiving, is essentially different

from location spoofing due to uncertainty about the inherent characteristics of this information due to the positioning technique of the mobile devices. Most Smartphone's simultaneously support a variety of positioning methods, including GPS (or other navigation satellite systems, such as Beidou of China, Galileo of E.U., and GLONASS of Russia), WLAN (Wireless Local Area Network) tracking system, Cell-ID positioning systems, and even Wi-Fi, Bluetooth or RFID based indoor positioning systems. These systems are associated with various levels of uncertainty. According to Watzdorf and Michahelles (2010), the positioning accuracy of GPS is between five and 10 meters, WLAN is sufficiently accurate within 30 to 50 meters. Cell-ID based positioning requires mobile network coverage, but it becomes less accurate within several hundred meters (Küpper 2005). Accuracy increases greatly with the Wi-Fi based Positioning System (WPS), which combines nearby Wi-Fi locations to provide accuracy similar to GPS (Lassabe et al. 2009). However, it only functions indoors or at least within Wi-Fi coverage areas. Though a hybrid solution can greatly improve the accuracy within three meters (Shaner 2013), it cannot eliminate locational uncertainty.

Although some uncertainty is inevitable, its impact can be mitigated by a variety of means. For example, in Figure 53, to make the audience aware of locational uncertainty, Apple map uses a lighter transparent circle to represent an uncertain area rather than simply

mapping a fixed spot. In addition, Goodchild (1998) has mentioned several other means, including modeling the locational information via a fuzzy set (Altman 1994, Robinson 2003), describing detailed metadata (Comber et al. 2006), or resorting to cartographic techniques, such as blurring the feature, or fading the colors to distinguish the uncertain area.

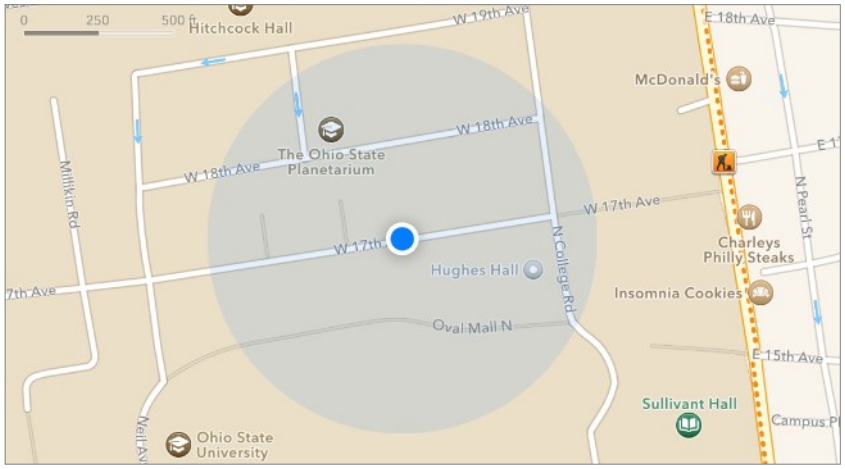

Figure 53. Representation of locational uncertainty

- Software design dilemma

To ensure that the locational information precisely represents the ground truth, the developers of social media apps must design an effective mechanism to control the quality of the locational information. But no matter how complicated or advanced the mechanism, there are some inherent dilemmas which make it impossible for the locational information to precisely represent the ground truth and consequently confuse the audience.

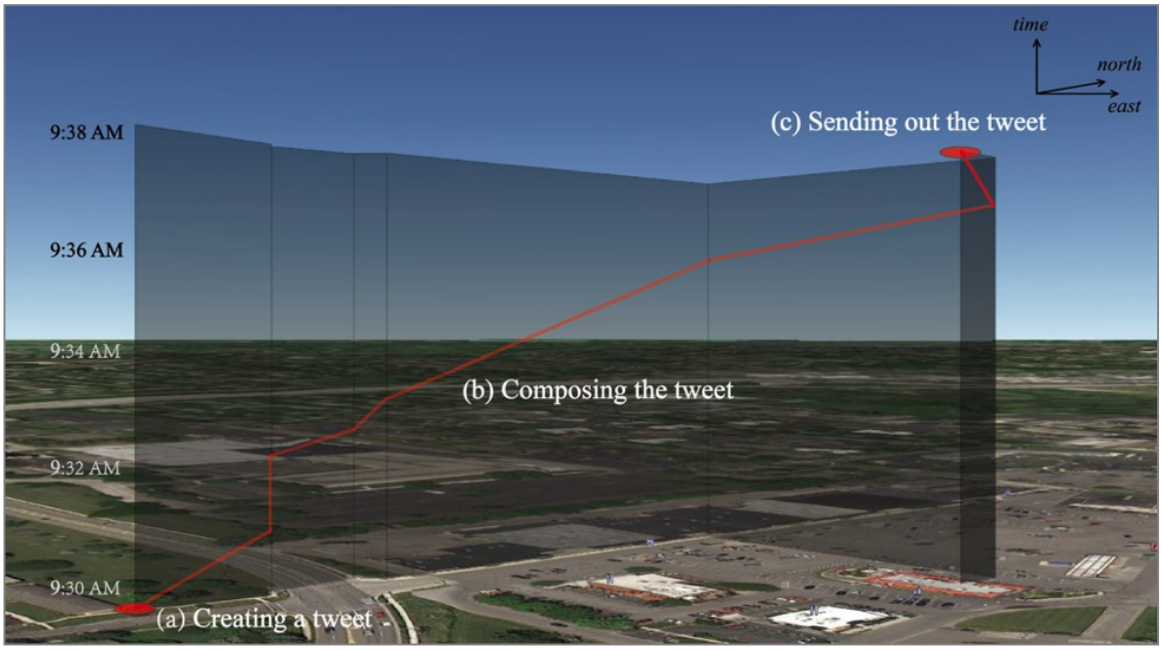

Figure 54. Analysis of the posting process of a tweet

To unveil the dilemma within app design, I investigated the entire process of "tweeting", which consists of three consecutive steps - creating, composing and finally sending out a tweet. For example, Figure 54 demonstrates the process of posting a geo-tagged tweet while walking. To start the process, a user creates a tweet at 9:30 AM. Simultaneously, the mobile app acquires the locational information to process it as a geo-tag. While walking, the user is composing the content of this tweet, but the geo-tag is not synchronized to the user's location. When the user is finished writing the tweet, he or she presses the "Tweet" button. In a second or two, the tweet is sent to the server at 9:38 AM, and then it becomes publicly accessible. At this point, the geo-tag is still pointing to the very place where the tweet was created at 9:30 AM, whereas the "created at"/timestamp property indicates the

very moment of sending out the tweet at 9:38 AM. In this sense, the spatial and temporal properties of the given tweet actually do not represent the same space-time event; thus, from a strict conceptual perspective, this locational information does not accurately represent the geographic ground truth.

Of course, social media developer teams are capable of taking various actions to handle this issue, but honestly, any expedient solution might inevitably result in new issues. Developers must ask themselves such questions as, which scenario more accurately represents the timestamp property of a tweeting activity, a single moment or a longer period of time? Or is it more reasonable to use a point feature rather than a polyline to represent the spatial property of the entire process of tweeting? In fact, it is difficult and even impossible to propose a best-fit solution under these difficult circumstances. Therefore, to minimize potential problems, software developers are obliged to warn online users of this design dilemma lying in any location based functions.

- Bug or computer malfunction

Motivation is one defining dimension of location spoofing, which originates from the consciousness of human beings or the collective will of societal organizations. Referring back to the four levels of mens rea (see Section 2.1.5), the location spoofer shall be a reasonable person. Otherwise, the user who created a piece of fake locational information is not a location spoofer. Indeed, it is quite controversial to extend the research scope of

location spoofer to embrace non-humans such as computers, robots or any other type of machines, though they may also be involved in spoofing.

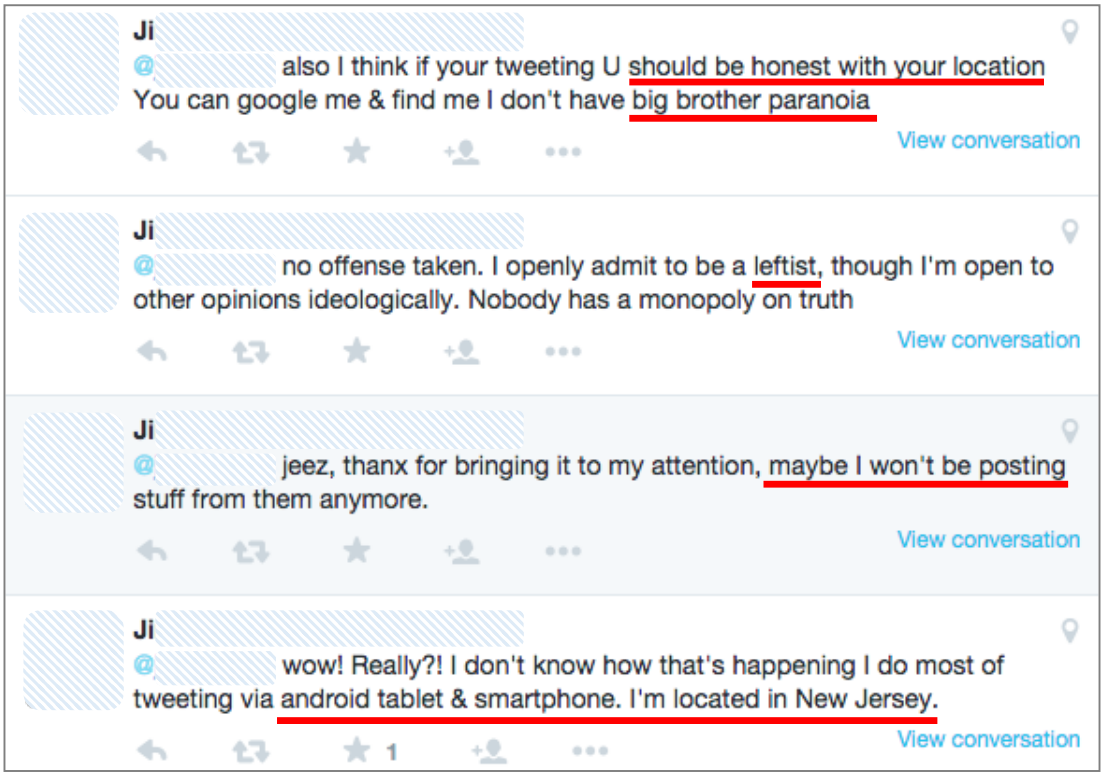
Ji
@ also I think if your tweeting U should be honest with your location You can google me & find me I don't have big brother paranoia
View conversation
Ji
@ no offense taken. I openly admit to be a leftist, though I'm open to other opinions ideologically. Nobody has a monopoly on truth
View conversation
Ji
@ jeez, thanx for bringing it to my attention, maybe I won't be posting stuff from them anymore.
View conversation
Ji
@ wow! Really?! I don't know how that's happening I do most of tweeting via android tablet & smartphone. I'm located in New Jersey.
1
View conversation

Figure 55. Interview text with a leftist from New Jersey

The proposed detecting approach has already found a fake geo-tag at the North Pole possibly caused by a malfunctioning Android tablet. Figure 55 shows part of the interview text with the author of the fake geo-tag. He is a die-hard leftist living in Parsippany, New Jersey, who was very active in his local labor union. From my observations, most of his geo-tags originated from New Jersey, but occasionally they appeared to come from the North Pole via an Android tablet. Due to the extremely short time interval between sending

two consecutive geo-tagged tweets, respectively from New Jersey and the North Pole, I was able to confirm that the geo-tag at the North Pole was a fake. However, surprisingly, this interviewee was unaware of how this spoofing phenomenon occurs, and emphasized that he had never done anything in particular with the locational configuration of the Twitter app and devices. Moreover, considering the political implications of geo-tags to draw attention from local people, it seems useless for a local leftist activist to geo-tag such a remote place as the North Pole. Hence, I could not identify any obvious spoofing motivation.

Still, it is problematic to attribute this fake geo-tag to the uncertainty or software design issue, because the extremely long offset is far beyond the uncertainty level of locational information, and it can be absolutely avoided by the quality controls of the Twitter official app for Android tablet. In addition, in the entire data sample, only this user was found to post tweets to the North Pole. Furthermore, regarding the raw data retrieved from API, the coordinate values of a geo-tag usually have some digits after the decimal point, but the fake geo-tag at the North Pole, In addition, the longitude is stated as exactly 0 and the latitude is set at 90. Therefore, I believe that this fake geo-tag at the North Pole originated from some bugs or tablet malfunctions. Certainly, computers and robots are capable of modifying information either unintentionally (e.g., bug, malfunction, etc.) or intentionally (e.g., computer virus). However, it is way beyond the scope of this research to discuss whether

non-humans can possess self-motivation, or whether human beings essentially have complete control of them. These topics delve into the ethics of computers or robots per se. However, practically, we should not ignore the growing multi-faceted implications of computer/robots to our society, especially how they are able to reform the human perception of space-time.

4.4.3 Social implications of location spoofing

Intuitively, location spoofing may cause negative effects in society, which seem to coincide with its reversal effects under McLuhan's tetradic framework. Several scholars have examined this reversal phenomenon and have given it different names, including "Marx's alienation" (Ollman 1977) to "technopoly" (Postman 1993). Today, the popularity of geospatial big data, especially the overloaded multimedia content (e.g. texts, images, audios, videos) of social media feeds, has fostered hyper-real illusions/spoofings about the geographic environment around us. Indeed, location spoofing has inundated us with a flood of fake GPS signals, lies, deceptive advertisements, spam and cyber-espionage. For example, referring to two previous examples, the fiancée, mentioned at the beginning of my dissertation, absolutely trusted that Bob was working hard in the lab. Additionally, the captain truly believed that the yacht had veered off the planed course. When this hyper-reality developed to extreme, it may flip into its reverse (from hyper-real to

hyper-fake) to deceive human beings. This spoofing phenomenon can result in unpredictable social consequences. We can imagine, if the fiancée discovered Bob had secretly deceived her by going to a bar, their wedding might have been postponed or even cancelled. Also, if the captain had altered the ship's course based on the erroneous message, the yacht might have wrecked into a reef. Undoubtedly, if we humans are not aware of this, Heidegger's fears (1954) about any new technologies might come true. Or, we might inevitably lock ourselves into the "iron cage" (Weber 2002) forged by that new technology. When we unthinkingly turn to geospatial big data for important messages, rather than considering their creditability, we might be fooled by these messages per se. Thus, seemingly, location spoofing can only result in negative social implications.

Instead of merely surrendering to these implications, McLuhan warned that these laws were mutually constructed, and four elements were the foundation of his laws of media. Only then can an integral awareness be developed. Therefore, while the reversal effects deceive us and generate negative implications, location spoofing simultaneously enhances the capability to virtually visit almost anywhere in the world. This enhancement is consistent with McLuhan's argument: "All media are extensions of some human faculty". It not only allows us to record visited places and visualize individual memory space, but also provides a means by which to struggle against the cyber-surveillance of big (as well as little) brother and thereby protect individual geo-privacy. Location spoofing can make

dominant location based services invalid and obsolete. An extreme but closely related case concerns the missing Malaysian Airline flight 370. Because its location was concealed and no one was able to discover it, this missing aircraft could be viewed as one of the most successful location spoofing cases in history. Ironically, in modern society, we claim to have various kinds of powerful and accurate location-based positioning and tracking systems (e.g., global navigation satellite systems, inertial reference systems [IRS], airplane ground tracking stations, etc.), we still are unable to locate a simple airplane. At the same time, location spoofing may also bring back Hollywood's geopiracy techniques as well as the military's deception in cyberspace. In most major social media platforms, online companies may fake their locations to distant areas, where they claim to provide services to the locals, thereby ignoring actual nearby potential customers. Today, cyber warfare, similar to other forms of real war, has taken advantage of location spoofing as a major type of military deception. Through multiple means (e.g., Tor, IP proxy, Fake location apps, etc.), cyber commands can fake or hide whereabouts before implementing any attacks.

By linking McLuhan's laws of media to the study of location spoofing, we can achieve a more holistic perspective about the potential impacts of it on society, and realize that its social implications have already transcended the dichotomous division of either negative or positive. Instead, they are always dialectical.

4.5 Summary

Relying on a hybrid methodology, this dissertation implemented a quantitative empirical method for detecting locational inconsistencies and a qualitative investigation on spoofing motivations and its social implications. Specifically, by linking location spoofing to Hägerstrand's time geography framework as well as Bayesian statistics, I developed a quantitative approach for detecting location spoofing in social media. After applying this approach to millions of geo-tagged tweets, I noticed that this approach was particularly effective for detecting certain types of location spoofing. In general, the detected location spoofing accounts for a trivial portion, approximately 0.22 percent of the entire sample, and this spoofing phenomenon seemingly had little overall impact on the data quality of social media feeds. However compared to the normal locational information, spoofing does have unique spatio-temporal dynamics, such as its special temporal preferences for performing spoofing and spatial distributions on different scales. More importantly, the social media feeds from various sources/platforms have been impacted by different degrees by location spoofing. For example, more evidence of spoofing can be observed on the third-party apps, such as Foursquare or Instagram, than on the official Twitter apps. Admittedly, this detecting approach does have some drawbacks. This approach would not be applicable for fake locational information, which is being continuously sent out at a fixed spot or goes beyond the human mobile constraints. However, in fact, it is not my

ultimate goal to develop a versatile approach for detecting all kinds of location spoofing in social media - an extremely challenging (if not entirely impossible) task. This proposed approach is still of significance in terms of automatically screening large-scale geo-tagged social media feeds and suggesting possible directions for further improvements, such as integrating with more specific geographic constraints in space and time, or utilizing the social hierarchical, crowdsourcing or linguistic identification capabilities.

After uncovering the fake locational information, I utilized a qualitative investigation to unveil the hidden spoofing motivations and other possible generative mechanisms. In addition, by viewing location spoofing in terms of McLuhan's laws of media, I have provided a holistic understanding about the broader impacts of this phenomenon on society. Specifically, while location spoofing has inundated us with location-based deceptive advertising, spam and cyber espionage, it also can potentially protect us against the location-based surveillance from big (as well as little) bother. Today, the geospatial techniques for location spoofing are not exclusively controlled by the government/military agencies or big corporations. An increasing number of users can easily spoof their locational information through "fake location" apps or GPS jammers/spoofers for multiple purposes, such as hiding their whereabouts, avoiding tracking, recording visited places, etc. Moreover, for big corporations that provide social media services, location spoofing is an effective method for protecting individual geo-privacy as well as an advertising strategy to

stimulate local business. In a much broader sense, for government/military agencies, location spoofing has become a necessary cyber-weapon for conducting espionage or preventing online tracking. Overall, this discussion reveals the multiple dimensions of this spoofing phenomenon and suggests ways to directly tackle this thorny issue in location-based social media or even in everyday life.

Chapter 5: Conclusions and Future Research

This dissertation focused on an often-ignored geographic phenomenon -"location spoofing" for the first time from a GIScience and geography perspective. This phenomenon, though rapidly emerging in location-based social media, actually has existed in different forms throughout human history. By employing both quantitative and qualitative analyses of various relevant spoofing activities, this dissertation filled in the lacunae in the GIScience and geography literature by examining the role of location spoofing in social media. In addition, this dissertation also discussed its broader implications for society, and called for reactions from GIScientists and the public at large to solve this thorny issue and be aware of its potential consequences. To sum up, the rest of this chapter wraps up this dissertation with a summary of the entire dissertation and concluding remarks, and discusses several directions for future research.

5.1 Summary and concluding remarks

This dissertation defined location spoofing as an intentional act of masking one's true location, and examined it as an information transmission process with several complex elements (e.g., location spoofer, fake locational information, spoofed audience, etc.). I

argued that location spoofing is essentially driven by individual or organizational motivations and manifests itself as a locational inconsistency between the geographic ground truth and the observed locational information.

Rather than siding with either positivism or interpretivism, I adopted a critical realist angle to investigate location spoofing. This perspective enabled me to holistically conceptualize this spoofing phenomenon, interpret hidden motivations, develop a practical detecting methodology, and uncover the broader social implications of spoofing. More specifically, I developed a quantitative geographic approach to detect the fake locational information. This kind of information is essential in documenting or representing locational inconsistency. Moreover, to unveil the complex motivations and other generative mechanisms of location spoofing, I relied on qualitative approaches, such as online observations, interviews, and case studies.

The proposed detecting approach is comprised of two successive steps: (a) filter: identifying a fake space-time path, of which at least one starting or ending point is falsified, (b) refine: determining the fake locational information from a fake path. After applying this approach to millions of geo-tagged tweets, I tested its utility in detecting certain types of location spoofing. Overall, the detected location spoofing only accounted for an insignificant portion of the entire sample. Due to this small ratio, this spoofing phenomenon may have little impact on the overall data quality. Even so, compared to

normal locational information, location spoofing does represent some unique spatial-temporal dynamics, and these features vary between social media platforms/sources (e.g., official clients, third-party apps, etc.). Admittedly, this detecting approach has some limitations. For fake locational information that is being continuously sent out at a fixed spot or does not violate human mobile constraints, this approach is not applicable. However, in fact, it was never my ultimate goal to develop a versatile approach for detecting all kinds of location spoofing in social media - an extremely challenging (if not entirely impossible) task. This proposed approach is still of great value in terms of automatically screening large-scale geo-tagged social media feeds and suggesting possible directions for further study, such as how to integrate with more targeted geographic constraints in space and time, or how to more efficiently utilize social hierarchical, crowdsourcing or linguistic identification techniques.

After uncovering the spoofed locational information, I conducted a qualitative investigation to explore the hidden spoofing motivations and/or other generative mechanisms. I found that location spoofing has been (ab)used by individual users, social media providers, and government/military agencies, and big corporations (e.g., social media providers) for multiple purposes. By linking location spoofing to McLuhan's laws of media, I was able to analyze its multi-faceted impacts on our daily lives, social developments, and even national security. In conclusion, this dissertation makes the

following contributions to the fields of GIScience and geography:

(1) Theoretically, this dissertation takes on the first systematic research of the spoofing phenomenon in geographic and GIScience context. By utilizing a critical realist perspective, I carried out an in-depth study on the multi-facets of location spoofing, which include defining relevant vocabularies, analyzing its information transmission and constituent elements, critically investigating the underlying locational inconsistencies as well as hidden spoofing motivations, and discussing social implications.

(2) Methodologically, I employed a hybrid methodology to embrace a quantitative empirical study on detecting locational inconsistency and a qualitative investigation on profiling the spoofing motivations as well as discussing the social implications. Specifically, this study further improved time geography with Bayesian statistics. Compared to the binary or frequentist time geographic approach, a Bayesian time geographic approach is capable of integrating the strengths of individual behavior (prior beliefs) in light of new evidence about collective human behaviors (as the posteriori). Moreover, a data-driven approach was adopted to formulate human appearance probability, which envisions the potentials of geospatial big data in dealing with geographic issues. Furthermore, the qualitative investigation on spoofing motivation provides a solid, practical foundation for further studies on users' motives and reactions to geospatial techniques. In addition, McLuhan's laws of media may shed light on the potential impacts

of new geospatial techniques on society.

(3) Computationally, to make the expensive and difficult high-performance computing more accessible for academic research, this dissertation, for the first time in GIScience, proposes a means to assemble a crawling cluster with open source chips. This cluster can be effectively utilized for harvesting a large-scale data set. It also possesses the potential for implementing specific geospatial computations.

Overall, this dissertation, which represents a synergy that includes cognitive, computational and social/behavioral sciences, will shed new light on multiple theoretical perspectives developed in GIScience studies, ranging from uncertainty and data quality assessment to spatial-temporal and social dimensions of online human behavior. The proposed detecting method has many applications, such as improving location-based services, protecting individual geo-privacy (e.g., location spoofing warning system), and sustaining cyber security (e.g., defending cyber-attacks by hiding online locations). More importantly, I hope to encourage social media users, service providers and regulators to directly tackle this rapidly emerging topic in social media and other forms of geospatial big data. Only then can we holistically understand various types of location spoofing and their generative mechanisms. Though still at a preliminary stage, I hope this study provides awareness about the multiple dimensions of this spoofing phenomenon in the era of big data.

5.2 Future research

In terms of future study, I am quite interested in the following topics as immediate follow up studies:

5.2.1 Integrating with other detecting approaches

To improve detection, I would like to learn more about and utilize more geographic knowledge and other alternative detecting approaches.

Currently, the proposed detecting approach is based on Hägerstrand's time geography, which is more like a bottom-up restriction of behavior in space-time. Other than individual constraints, normal locational information on social media should also follow some macro level controls, such as Tobler's first law of geography, central place theory, or the rejection of complete spatial randomness (CSR). For example, if the spatial pattern of all geo-tags does not follow CSR, it is highly possible that some of the geo-tags from this user are spoofed. However, the drawback of this macro level control is obvious - it is impossible to confirm which one of the geo-tags is spoofed, also this approach relies on a prerequisite. That is, there should be plenty of geo-tags from this given user to generate a valid spatial pattern. Otherwise, this approach becomes ineffective.

Other than the above listed geographic approaches, I am interested in proposing a

crowdsourcing approach. Indeed, collective wisdom can help us better understand impossible or inconsistent geographic phenomena. A crowdsourcing approach usually requires a substantial number of observations about suspected location spoofing. Notably, even though there are undeniable merits of crowdsourcing, the collective knowledge from social media users are by no means synonymous with the truth. Therefore, scholars should question any collective decision.

Moreover, the proposed detecting approach uses a dichotomous strategy to describe a piece of locational information, which is either spoofed or not. However, in reality, it is difficult to make such a determination. Instead of employing this dichotomous strategy, the proposed detecting approach might better measure the confidence level, or how likely a user is to believe each piece of spoofed locational information. In statistics, the confident level represents how often the true percentage of the total population who would pick an answer lies within the confidence interval. The 95% confidence level means one can be 95% certain, while the 99% confidence level means one can be 99% certain. Accordingly, by utilizing the confident level, the proposed approach can avoid describing location spoofing with the simple dichotomous terms – spoofing or not spoofing. Instead, the ability to calculate a percentage of the confident level provides a more accurate statistical and scientific interpretation.

Overall, to improve detection, it is necessary to formulate an integrated method for

modeling more geographic knowledge and combining both automatic and manual screening is a necessity. In addition, from a statistical standpoint, the confidence level allows us to more scientifically interpret any individual case of location spoofing.

5.2.2 Cluster based spoofing detection

Unlike conventional computing techniques, cluster computing increases capabilities based on parallel mechanisms. In the proposed detecting approach, cluster computing is merely used in data collection, but in fact, it can improve multiple aspects of the proposed approach. For example a cluster can fundamentally enhance efficiency when the proposed approach is employed to build a location spoofing warning system. Computing human appearance probability, for example, made use of geo-tags for approximately three days. A three-day period may be enough to accurately represent the temporal preferences of a day, but it falls short for uncovering the preferences in a week and even the temporal distinctions between weekdays or weekends. For this reason, a larger data sample with a longer period of geo-tags is required. To harvest such a data set, a cluster will provide high-performances than traditional workstations. Moreover, to determine the threshold of tweet density, I will use an effective approach to count geo-tagged tweets within a fixed-size window. If running in a higher performance environment, the proposed method can switch to a more sophisticated and scientific spatial interpolation algorithm (e.g.,

kernel density) to derive this variable.

More importantly, the state-of-the-art GeoComputation allows scholars to establish such a cluster. Indeed, for geospatial big data storage, there are NOSQL data storage (e.g., MongoDB, HDFS) and relational databases (e.g., PostgreSQL, MySQL), for spatial interpolations with parallel computing functions, and ready-to-use spatial libraries (e.g., PySAL, geomesa). In addition to the software, the burgeoning open-source hardware movement has made the digital parts necessary for assembling a cluster much less expensive and readily available. Thus, thanks to these software and hardware developments, a cluster is more capable of detecting location spoofing in a timely manner, and we can develop more flexible counter-measures to deal with this issue.

5.2.3 Further critical inquiries

Although McLuhan's framework can help us better understand the multi-faceted nature of location spoofing, its solution for dealing with it is highly controversial. McLuhan once stated, "One should charge straight ahead and kick them (electronic media)", and then the media will "respond beautifully" and "soon become servants rather than masters" (McLuhan and Zingrone 1995, 267-268). This argument was no more than a Luddite-type reaction against location spoofing. At the other extreme are those who would like to blindly adopt any "new technology", and believe that it will eventually trump human will (Carey

2007). It is human nature not to allow location spoofing to dictate individual behaviors or dominate social developments. But while dealing with any aspect (e.g., location spoofing) of society, we inevitably confront a prevailing dilemma about self-reference - the approach we use to study society is actually a part of the society itself. Perhaps, to transcend this fractal-like status quo, we must clearly alter and continuously re-alter our social goals to examine and re-examine location spoofing under these overarching social goals.

To achieve this goal, my further research will explore a possible solution, that is, to shift from instrumental to commutative rationality (Habermas 1984). This method will encourage us to perceive location spoofing as an interaction with other social factors, including people, society, the natural environment, and even those controversial non-human objects (e.g., machines, robots, etc.). During the course of an interaction, profound social and psychological effects can help the participants to adjust their goals. When the goals are integrated into an organic whole, the overall social goal will evolve, as will our perception of location spoofing or any other forms of new geospatial technology.

References

Adams, P. C., 2009. *Geographies of Media and Communication.* Oxford, UK: John Wiley & Sons.

Altman, D. 1994. Research Article. Fuzzy Set Theoretic Approaches for Handling Imprecision in Spatial Analysis. *International Journal of Geographical Information Systems,* 8(3), 271-289.

Ansari, A. 2012. The Role of Social Media in Iran's Green Movement (2009-2012). *Global Media Journal-Australian Edition,* 6(2), 1-6.

Archer, M.*, et al.*, 2013. *Critical Realism: Essential Readings.* London, UK: Routledge.

Batty, M.*, et al.* 2012. Smart Cities of the Future. *The European Physical Journal Special Topics,* 214(1), 481-518.

BBC, 2014. *Us Confirms It Made 'Cuban Twitter'* [online]. Available from: http://www.bbc.com/news/world-us-canada-26876801 [Accessed May 24 2015].

Bhaskar, R. A., 1975. *A Realist Theory of Science.* York, UK: Alma Book Company.

Bhaskar, R. A., 1979. *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences.* 3 ed. London, UK: Routledge.

Bhaskar, R. A., 1993. *Dialectic: The Pulse of Freedom.* London, UK: Verso.

Bhatnagar, B., 2014. *How to Spoof Your Iphone Location on iOS 7* [online]. Available from: http://www.iphonefaq.org/archives/973407 [Accessed May 24 2015].

Bolino, M. C. 1999. Citizenship and Impression Management: Good Soldiers or Good Actors? *Academy of Management Review,* 24(1), 82-98.

Bordia, P., Irmer, B. E. and Abusah, D. 2006. Differences in Sharing Knowledge Interpersonally and via Databases: The Role of Evaluation Apprehension and Perceived Benefits. *European Journal of Work and Organizational Psychology,*

15(3), 262-280.

Bordogna, G.*, et al.* 2014. A Linguistic Decision Making Approach to Assess the Quality of Volunteer Geographic Information for Citizen Science. *Information Sciences,* 258, 312-327.

Boyd, D. and Crawford, K. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society,* 15(5), 662-679.

Brewin, B., 2013. *GPS Jammers Could Knock out Signals in a Medium-Sized City* [online]. Available from: http://www.nextgov.com/defense/2013/09/gps-jammers-could-knock-out-signals-medium-sized-city/70582/ [Accessed May 24 2015].

Burton, G. and Dimbleby, R., 1988. *Between Ourselves: An Introduction to Interpersonal Communication.* Edward Arnold London.

Carey, J. W. 2007. A Short History of Journalism for Journalists: A Proposal and Essay. *The Harvard International Journal of Press/Politics,* 12(1), 3-16.

Castiglione, A., De Prisco, R. and De Santis, A., 2009. Do You Trust Your Phone? *In:* DI NOIA, T. AND BUCCAFURRI, F. eds. *E-Commerce and Web Technologies.* Berlin Heidelberg: Springer, 50-61.

Cctv, 2013. *微信色情调查 (Wechat Pornography Survey)* [online]. Available from: http://tv.cntv.cn/video/C22416/ee640e17f4b74e01a596ec239f2c50d9 [Accessed May 24 2015].

Chartrand, S., 2004. *New Way to Combat Online Piracy* [online]. NY Times. Available from: http://www.nytimes.com/2004/05/17/technology/17patent.html [Accessed May 24 2015].

Cheng, Z., Caverlee, J. and Lee, K., 2010. You Are Where You Tweet: A Content-Based Approach to Geo-Locating Twitter Users. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management.* Toronto, ON, Canada: ACM, 759-768.

Chung, D. 2001. Electromagnetic Interference Shielding Effectiveness of Carbon Materials. *Carbon,* 39(2), 279-285.

Coleman, D. J., Georgiadou, Y. and Labonte, J. 2009. Volunteered Geographic Information: The Nature and Motivation of Produsers. *International Journal of Spatial Data Infrastructures Research,* 4(1), 332-358.

Comber, A. J.*, et al.*, 2006. Using Metadata to Link Uncertainty and Data Quality Assessments. *Progress in Spatial Data Handling.* Springer, 279-292.

Crampton, J. W.*, et al.* 2013. Beyond the Geotag: Situating 'Big Data' and Leveraging the Potential of the GeoWeb. *Cartography and Geographic Information Science,* 40(2), 130-139.

Croitoru, A.*, et al.* 2013. Geosocial Gauge: A System Prototype for Knowledge Discovery from Social Media. *International Journal of Geographical Information Science,* 27(12), 2483-2508.

De Born, Z., 2014. *The Photo Stream of Zilla's Facebook* [online]. Available from: https://www.facebook.com/zillavandenborn/photos_stream?tab=photos_stream [Accessed May 24th 2015].

Dean, J. and Ghemawat, S. 2008. Mapreduce: Simplified Data Processing on Large Clusters. *Communications of the ACM,* 51(1), 107-113.

Dodge Jr, R. C., Carver, C. and Ferguson, A. J. 2007. Phishing for User Security Awareness. *Computers & Security,* 26(1), 73-80.

Downs, J. A., 2010. Time-Geographic Density Estimation for Moving Point Objects. *Geographic Information Science.* Springer, 16-26.

Drager, D., 2011. *How to Disable or Fake Your Location in Firefox, Internet Explorer & Chrome* [online]. Available from: http://www.makeuseof.com/tag/disable-fake-location-firefox-internet-explorer-chrome/ [Accessed May 24 2015].

Economists, 2013. *Out of Sight* [online]. Available from: http://www.economist.com/news/international/21582288-satellite-positioning-data-are-vitalbut-signal-surprisingly-easy-disrupt-out [Accessed May 24 2015].

Elwood, S. E., Goodchild, M. F. and Sui, D. Z. 2012. Researching the Geocrowd: Volunteered Geographic Information and Geographic Research. *Annals of the AAG,* 102(3), 571-590.

Esmaili, R., Naseri, F. and Esmaili, A. 2013. Quality Assessment of Volunteered Geographic Information. *American Journal of Geographic Information System,* 2(2), 19-26.

FCC 2012. Consumer Alert: Using or Importing Jammers Is Illegal. *Federal Communications Commission*.

Felten, E. W.*, et al.* 1997. Web Spoofing: An Internet Con Game. *Software World,* 28(2), 6-8.

Fill, C. 1999. Marketing Communications, Context, Contents and Strategies. 2 ed. New Jersey: Prentice Hall.

Fishman, D. A. 2006. Review and Criticism: Research Pioneer Tribute—Rethinking Marshall McLuhan: Reflections on a Media Theorist. *Journal of Broadcasting & Electronic Media,* 50(3), 567-574.

Flanagin, A. J. and Metzger, M. J. 2008. The Credibility of Volunteered Geographic Information. *GeoJournal,* 72(3-4), 137-148.

Ford, M., 2014. *#Cancelcastro: Why Is U.S. Policy toward Cuba So Absurd?* [online]. Available from: http://www.theatlantic.com/international/archive/2014/04/-cancelcastro-why-is-us-policy-toward-cuba-so-absurd/ [Accessed May 24 2015].

Fotheringham, A. S. 1981. Spatial Structure and Distance-Decay Parameters. *Annals of the Association of American Geographers,* 71(3), 425-436.

Galt, R. E. 2011. Counting and Mapping Community Supported Agriculture (CSA) in the United States and California: Contributions from Critical Cartography/GIS. *ACME: An International E-Journal for Critical Geographies,* 10(2), 131-162.

GCN, 2015. *CyberGIS: Infrastructure for Massive Geospatial Data, Processes* [online]. Available from: http://gcn.com/articles/2015/02/04/cybergis.aspx [Accessed May 25 2015].

Gibbons, G., 2013. *FCC Fines Operator of GPS Jammer That Affected Newark Airport GBAS* [online]. Available from: http://www.insidegnss.com/node/3676 [Accessed Feburary 3rd 2014].

Gilmer, J., 2015. *All the Geo-Tagged Locations of Uploaded Flickr Photos by Concentration* [online]. Birmingham, Alabama: Huebris. Available from: http://huebris.com/services/marketing/social-media-consulting/ [Accessed May 24 2015].

Goodchild, M. 1998. Uncertainty: The Achilles Heel of GIS. *Geo Info Systems,* 8(11), 50-52.

Goodchild, M. F. 1993. Data Models and Data Quality: Problems and Prospects. *Environmental Modeling with GIS*, 94-103.

Goodchild, M. F., Spatial Accuracy 2.0. ed. *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 2008, 1-7.

Goodchild, M. F. 2013. The Quality of Big (Geo) Data. *Dialogues in Human Geography,* 3(3), 280-284.

Goodchild, M. F.*, et al.* 1999. Introduction to the Varenius Project. *International Journal of Geographical Information Science,* 13(8), 731-745.

Goodchild, M. F. and Li, L. 2012. Assuring the Quality of Volunteered Geographic Information. *Spatial statistics,* 1, 110-120.

Graham, M. and Zook, M. 2011. Visualizing Global Cyberscapes: Mapping User-Generated Placemarks. *Journal of Urban Technology,* 18(1), 115-132.

Griffin, E., 2003. *A First Look at Communication Theory.* 4th ed. Boston: McGraw-Hill.

Guo, C. and Jiang, Y., 2014. *石河子莫索湾公安局抓获一名网上在逃人员 (an Outlaw Has Been Caught by Policemen of Mosuowan, Shihezi City)* [online]. Available from: http://www.ts.cn/shz/2014-02/27/content_9371683.htm [Accessed May 24 2015].

Gupta, A. K. and Govindarajan, V. 2000. Knowledge Flows within Multinational Corporations. *Strategic Management Journal,* 21(4), 473-493.

Habermas, J. 1984. The Theory of Communicative Action, Vol. I. *Boston: Beacon*.

Hägerstrand, T. 1970. What About People in Regional Science? *Papers in regional science,* 24(1), 7-24.

Hägerstrand, T. 1982. Diorama, Path and Project. *Tijdschrift voor Economische en Sociale Geografie,* 73(6), 323-329.

Hahn, E., A Linguistic Fallacy. ed. *Studies Presented to Joshua Whatmough on His Sixtieth Birthday*, 1957, 53-64.

Haklay, M.*, et al.*, 2014. *Crowdsourced Geographic Information Use in Government.* London, UK: Work Bank.

Harvey, A., Mose, A. and Burtch, A., 2013. *Anticipating the Future of DIY DNA Counter-Surveillance* [online]. Ahprojects. Available from: http://ahprojects.com/projects/dna-spoofing/ [Accessed May 24 2015].

Heidegger, M. 1954. The Question Concerning Technology. *Technology and Values:*

Essential readings, 99-113.

Henning, P. J., 2014. *Spoofing, a New Crime with a Catchy Name* [online]. NY Times. Available from: http://dealbook.nytimes.com/2014/10/06/a-new-crime-with-a-catchy-name-spoofing/ [Accessed May 24 2015].

Hill, K., 2014. *Stolen Nude Images Reveal Celebs' Location Information* [online]. Forbes. Available from: http://www.forbes.com/sites/kashmirhill/2014/09/02/leaked-nude-images-reveal-celebs-location-information/ [Accessed November 15th 2014].

Humphreys, T. E.*, et al.*, Assessing the Spoofing Threat: Development of a Portable GPS Civilian Spoofer. ed. *Proceedings of the ION GNSS International Technical Meeting of the Satellite Division*, 2008, 2314-2325.

Java, A.*, et al.*, 2009. Why We Twitter: An Analysis of a Microblogging Community. *Advances in Web Mining and Web Usage Analysis.* Springer, 118-138.

Kalantari, M.*, et al.* 2014. Geospatial Metadata 2.0–an Approach for Volunteered Geographic Information. *Computers, Environment and Urban Systems,* 35-48.

Kankanhalli, A., Tan, B. C. and Wei, K.-K. 2005. Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation. *Mis Quarterly*, 113-143.

Keach, S., 2014. *Instagram Now Has More Users Than Twitter* [online]. Available from: http://www.trustedreviews.com/news/instagram-now-has-more-users-than-twitter [Accessed May 24 2015].

Kitchin, R.*, et al.* 2013. Public Geographies through Social Media. *Dialogues in Human Geography,* 3(1), 56-72.

Koerner, B., 2014. *Inside the New Arms Race to Control Bandwidth on the Battlefield* [online]. Available from: http://www.wired.com/2014/02/spectrum-warfare/.

Kuklinski, A., 1987. Torsten Hägerstrand Laudatio. *GeoJournal* 14 (4): 503-510.

Küpper, A., 2005. Location-Based Services: Fundamentals and Operation. Chichester, England: Wiley.

Lacan, J., 1960. The Subversion of the Subject and the Dialectic of Desire in the Freudian Unconscious. *In:* KEENAN, D. K. ed. *Hegel and Contemporary Continental Philosophy.* Albany, NY: SUNY Press.

Lassabe, F.*, et al.* 2009. Indoor Wi-Fi Positioning: Techniques and Systems. *Annals of*

Telecommunications-annales des Télécommunications, 64(9-10), 651-664.

Lasswell, H. D. 1948. The Structure and Function of Communication in Society. *The Communication of Ideas,* 37, 215-228.

Leszczynski, A. 2009. Poststructuralism and Gis: Is There a 'Disconnect'? *Environment and planning. D, Society and space,* 27(4), 581-602.

Levinson, P., 1999. *Digital Mcluhan: A Guide to the Information Millennium.* London, UK: Routledge.

Loshin, D., 2014. *Understanding Big Data Quality for Maximum Information Usability* [online]. SAS White Paper Series. Available from: http://www.sas.com/ro_ro/whitepapers/understanding-big-data-quality-107113.html [Accessed December 6 2014].

Mandiant, 2013. *Exposes One of China's Cyber Espionage Units* [online]. Alexandria, VA: Mandiant. Available from: http://intelreport.mandiant.com/Mandiant_APT1_Report.pdf [Accessed June 1 2014].

Mayer, R. C. and Gavin, M. B. 2005. Trust in Management and Performance: Who Minds the Shop While the Employees Watch the Boss? *Academy of Management Journal,* 48(5), 874-888.

Mcluhan, E. and Zingrone, F., 1995. *Essential Mcluhan.* Toronto, Canada: House of Anansi.

Mcluhan, M., 1964. *Understanding Media: The Extensions of Man.* New York: The New American Library.

Mcluhan, M. and Mcluhan, E., 1988. *Laws of Media: The New Science.* University of Toronto Press Toronto.

Mcluhan, M. and Powers, B. R., 1992. *The Global Village: Transformations in World Life and Media in the 21st Century.* Oxford, UK: Oxford University Press.

Miller, H. J. 2005. A Measurement Theory for Time Geography. *Geographical Analysis,* 37(1), 17-45.

Miller, H. J., 2014. Time Geography and Space-Time Prism. *In:* CASTREE, N. ed. *The International Encyclopedia of Geography: People, the Earth, Environment, and Technology.* Hoboken, NJ: Wiley.

Mingers, J. 2004. Real-Izing Information Systems: Critical Realism as an Underpinning

Philosophy for Information Systems. *Information and Organization,* 14(2), 87-103.

Monmonier, M., 1996. *How to Lie with Maps.* Chicago, IL: University of Chicago Press.

Murray, B. H., 2003. *Defending the Brand: Aggressive Strategies for Protecting Your Brand in the Online Arena.* New York: Amacom.

Mutzabaugh, B., 2013. *Routehappy Unveils Report Card for in-Flight Wi-Fi* [online]. USA Today. Available from: http://www.usatoday.com/story/todayinthesky/2013/06/25/routehappy-unveils-report-card-for-in-flight-wi-fi/2454565/ [Accessed Feburary 3rd 2014].

O'Neill, B. 2011. Media Literacy in Ireland: From Protectionism to Participation. *Communication Technology and New Media Commons*.

Ollman, B. 1977. Marx's Vision of Communism a Reconstruction. *Critique: Journal of Socialist Theory,* 8(1), 4-41.

Pandove, K., Jindal, A. and Kumar, R. 2010. Email Spoofing. *International Journal of Computer Applications,* 5(1), 27-30.

Papacharissi, Z. 2009. The Virtual Geographies of Social Networks: A Comparative Analysis of Facebook, Linkedin and Asmallworld. *New Media & Society,* 11(1-2), 199-220.

Perkins, C. 2003. Cartography: Mapping Theory. *Progress in Human Geography,* 27(3), 341-351.

Postman, N., 1993. *Technopoly: The Surrender of Culture to Technology.* New York: Vintage.

Press, A., 2014. *Errant Taliban Tweet Claims Spokesman in Pakistan* [online]. Available from: http://bigstory.ap.org/article/e47a93e78a76499e8282ff3c314bb091/errant-taliban-tweet-claims-spokesman-pakistan [Accessed Feburary 3rd 2015].

Raymond, E. S., 2001. *The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary.* Sebastopol, CA: O'Reilly Media.

Riesenfeld, R., 2011. *Bayes' Theorem* [online]. Available from: http://www.eng.utah.edu/~cs5961/Resources/bayes.pdf [Accessed October, 17 2014].

Robinson, V. B. 2003. A Perspective on the Fundamentals of Fuzzy Sets and Their Use in Geographic Information Systems. *Transactions in GIS,* 7(1), 3-30.

Saussure, F. D., 1974, 2013. *Course in General Linguistics.* New York: Columbia University Press.

Savage, C., Wyatt, E. and Baker, P. 2013. US Confirms That It Gathers Online Data Overseas. *The New York Times,* 6, 1898-2020.

Sayer, A., 1985. Realism and Geography. *In:* JOHNSTON, R. J. ed. *The Future of Geography.* London, UK: Methuen.

Sayer, A., 1992. *Method in Social Science: A Realist Approach.* 2 ed. London, UK: Routledge.

Schramm, W. 1954. How Communication Works. *The Process and Effects of Mass Communication*, 3-26.

Schuurman, N. 2002. Reconciling Social Constructivism and Realism in GIS. *ACME: An International E-Journal for Critical Geographies,* 1(1), 73-90.

Scott, M., 2014. *Google Provides Details on 'Right to Be Forgotten' Requests in E.U.* [online]. New York Times. Available from: http://bits.blogs.nytimes.com/2014/10/09/google-provides-details-on-right-to-be-forgotten-requests/ [Accessed November, 21th 2014].

Shaner, J., 2013. *Smartphones, Tablets and GPS Accuracy* [online]. ArcGIS. Available from: http://blogs.esri.com/esri/arcgis/2013/07/15/smartphones-tablets-and-gps-accuracy/ [Accessed May 24 2015].

Smith, P. A.*, et al.* 2006. Is Trust Really Social Capital? Knowledge Sharing in Product Development Projects. *The Learning Organization,* 13(6), 594-605.

Song, Y. and Miller, H. J. 2014. Simulating Visit Probability Distributions within Planar Space-Time Prisms. *International Journal of Geographical Information Science,* 28(1), 104 -125.

Southorn, G., 2015. *Easter Eggs in Maps Create Scrambled Destinations* [online]. Available from: http://360.here.com/2015/03/30/easter-eggs-maps-create-scrambled-destinations/ [Accessed May 24 2015].

Stasser, G. and Titus, W. 2003. Hidden Profiles: A Brief History. *Psychological Inquiry,* 14(3-4), 304-313.

Stefanidis, A., Crooks, A. and Radzikowski, J. 2013. Harvesting Ambient Geospatial

Information from Social Media Feeds. *GeoJournal,* 78(2), 319-338.

Sui, D. Z. 2014. Opportunities and Impediments for Open GIS. *Transactions in GIS,* 18(1), 1-24.

Sui, D. Z. and Goodchild, M. F. 2003. A Tetradic Analysis of GIS and Society Using Mcluhan's Law of the Media. *The Canadian Geographer/Le Géographe Canadien,* 47(1), 5-17.

Sui, D. Z. and Goodchild, M. F. 2011. The Convergence of Gis and Social Media: Challenges for Giscience. *International Journal of Geographical Information Science,* 25(11), 1737-1748.

Tanase, M., 2003. *IP Spoofing: An Introduction* [online]. Available from: http://66.14.166.45/sf_whitepapers/tcpip/IP Spoofing - An Introduction.pdf [Accessed May 24 2015].

Theall, D., 2001. *Virtual Marshall Mcluhan.* Montreal, Canada: McGill-Queen's Press-MQUP.

Twitter, 2011. *The Engineering Behind Twitter's New Search Experience* [online]. Available from: https://blog.twitter.com/2011/engineering-behind-twitter%E2%80%99s-new-search-experience [Accessed May 24 2015].

Vogel, J. H*., et al.* 2007. Geopiracy as an Emerging Issue in Intellectual Property Rights: The Rationale for Leadership by Small States. *Tul. Envtl. LJ,* 21, 391-406.

Wainwright, J., 2012. *Geopiracy: Oaxaca, Militant Empiricism, and Geographical Thought.* New York, NY: Palgrave Macmillan.

Watzdorf, S. V. and Michahelles, F., 2010. Accuracy of Positioning Data on Smartphones. *Proceedings of the 3rd International Workshop on Location and the Web.* Tokyo, Japan: ACM, 1-4.

Weber, M., 2002. *The Protestant Ethic and the" Spirit" of Capitalism and Other Writings.* London, UK: Penguin Classics.

Wechsler, H*., et al.*, 1962. *Model Penal Code.* Philadelphia, PA: ALI (American Law Institute).

Willis, A., 2014. *Google Maps Easter Eggs Revealed: Hitch a Ride on the Loch Ness Monster or a Dragon in Wales* [online]. Available from: http://metro.co.uk/2014/06/03/google-maps-easter-eggs-revealed-hitch-a-ride-on-t

he-loch-ness-monster-or-a-dragon-in-wales-4748710/ [Accessed May 24 2015].

Wilson, M. W. 2012. Location-Based Services, Conspicuous Mobility, and the Location-Aware Future. *Geoforum,* 43(6), 1266-1275.

Winter, S. and Yin, Z.-C. 2010. Directed Movements in Probabilistic Time Geography. *International Journal of Geographical Information Science,* 24(9), 1349-1365.

Wood, J. 2008. Filter and Refine Strategy. *Encyclopedia of GIS*. New York, NY: Springer, 320.

Woollaston, V., 2015. *Taking the P***!! Google Maps Shows Android Urinating on Apple's Logo When Users Navigate to Pakistan* [online]. Available from: http://www.dailymail.co.uk/sciencetech/article-3053903/Taking-p-Google-Maps-shows-Android-URINATING-Apple-s-logo-users-navigate-Pakistan.html [Accessed May 24 2015].

Zaragoza, S., 2013. *Spoofing a Superyacht at Sea* [online]. Available from: http://www.utexas.edu/know/2013/07/30/spoofing-a-superyacht-at-sea/ [Accessed May 24 2015].

Zhang, J. and Goodchild, M. F., 2002. *Uncertainty in Geographical Information.* Boca Raton, FL: CRC press.

Zumalt, E., 2013. *Spoofing a Superyacht at Sea* [online]. Available from: http://news.utexas.edu/2013/07/30/spoofing-a-superyacht-at-sea [Accessed May 24th 2015].